

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2004, 19:22:54 ; Search time 177 Seconds
(without alignments)
4486.776 Million cell updates/sec

Title: US-09-522-753-5

Perfect score: 13215

Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEPKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

SPTREMBL_25:**

1: sp_archaea:**

2: sp_bacteria:**

3: sp_fungi:**

4: sp_human:**

5: sp_invertebrate:**

6: sp_mammal:**

7: sp_mhc:**

8: sp_organelle:**

9: sp_phase:**

10: sp_plant:**

11: sp_rodent:**

12: sp_virus:**

13: sp_vertebrate:**

14: sp_unclassified:**

15: sp_rvirus:**

16: sp_bacteriap:**

17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4149	31.4	1998	13	Q8QFM0
2	4146	31.4	1988	13	Q8QFL9
3	4043	30.6	2343	4	Q7Z516
4	4009.5	30.3	2498	13	Q8QG78
5	3178	24.0	606	4	Q9BT61
6	3012	22.8	631	11	Q80ZV7
7	2184	16.5	914	4	Q86YX0
8	2155	16.3	527	13	Q7SVR4
9	2091.5	15.8	1724	11	Q8CHH6
10	1839	13.9	1399	11	Q8VDE8
11	1687	12.8	550	4	Q86W52
12	1272	9.6	1047	4	Q9NS20
13	1172.5	8.9	959	4	Q86YX1
14	1077	8.1	3604	5	Q9VYX0
15	1075	8.1	3469	5	Q9U412
16	827.5	6.3	302	13	Q7ZUE9

17	650	4.9	453	4	Q86YY2	Q86yy2 homo sapien
18	633.5	4.8	291	11	Q8BK32	Q8bk32 mus musculus
19	582.5	4.4	17352	5	Q95YM2	Q95ym2 procambarus
20	577.5	4.4	5412	5	Q9W596	Q9w596 drosophila
21	574	4.3	5327	5	Q76891	Q76891 drosophila
22	555	4.2	1966	5	Q81QA6	Q81qa6 drosophila
23	553.5	4.2	1988	5	Q86BH2	Q86bh2 drosophila
24	552.5	4.2	2406	4	Q9BZS0	Q9bzso homo sapien
25	552	4.2	1966	5	Q9NHX6	Q9nhx6 drosophila
26	552	4.2	1985	5	Q8T9N4	Q8t9n4 drosophila
27	551.5	4.2	1985	5	Q9VSK5	Q9vsk5 drosophila
28	546.5	4.1	2157	4	Q9S875	Q9s875 homo sapien
29	543.5	4.1	2592	3	Q9P3J0	Q9p3j0 neurospora
30	536.5	4.1	2414	4	Q9HCL7	Q9hcl7 homo sapien
31	531.5	4.0	2157	4	Q96QC6	Q96qc6 homo sapien
32	522.5	4.0	2752	4	Q9UQ35	Q9uq35 homo sapien
33	521	3.9	2158	11	Q7TSC1	Q7tsc1 mus musculus
34	521	3.9	2607	11	Q8BT18	Q8bt18 mus musculus
35	520	3.9	2187	11	P70670	P70670 mus musculus
36	519	3.9	3536	5	Q9VZ30	Q9vz30 drosophila
37	517.5	3.9	2157	11	Q9Z1R1	Q9z1r1 mus musculus
38	513.5	3.9	34350	4	Q8WZ42	Q8wz42 homo sapien
39	508	3.8	5027	11	Q63460	Q63460 rattus norv
40	504	3.8	3942	11	Q88737	Q88737 mus musculus
41	503.5	3.8	7962	4	Q10465	Q10465 homo sapien
42	503	3.8	2344	5	Q9N3Y8	Q9n3y8 caenorhabdi
43	502	3.8	2465	5	Q81Q71	Q81q71 drosophila
44	498	3.8	3938	11	Q88778	Q88778 rattus norv
45	493.5	3.7	4322	10	Q7XXN1	Q7xxn1 oryza sativ
46	478.5	3.6	2296	4	Q9UHA8	Q9uha8 homo sapien
47	478	3.6	4900	5	Q9N541	Q9n541 caenorhabdi
48	477	3.6	2282	11	Q61479	Q61479 mus musculus
49	476.5	3.6	3851	4	Q43161	Q43161 homo sapien
50	476.5	3.6	3926	4	Q9UPA5	Q9upa5 homo sapien
51	475	3.6	3241	5	Q7YZH1	Q7yzh1 drosophila
52	473	3.6	3201	5	Q9W0U2	Q9w0u2 drosophila
53	469.5	3.6	2701	4	Q9Y520	Q9y520 homo sapien
54	459.5	3.5	2219	4	Q9C0A3	Q9c0a3 homo sapien
55	458.5	3.5	3111	5	Q9VH10	Q9vh10 drosophila
56	452.5	3.4	1566	4	Q9P2R6	Q9p2r6 homo sapien
57	452.5	3.4	2171	6	Q7YR40	Q7yr40 pan troglod
58	452	3.4	2971	4	Q9Y5L9	Q9y5l9 homo sapien
59	450.5	3.4	2296	4	Q9Y3S1	Q9y3s1 homo sapien
60	450.5	3.4	2768	5	Q9VC00	Q9vc00 drosophila
61	448.5	3.4	4969	11	Q8CF91	Q8cf91 mus musculus
62	448.5	3.4	5165	11	Q8CF92	Q8cf92 mus musculus
63	446	3.4	2948	4	Q86W66	Q86w66 homo sapien
64	445	3.4	18519	5	Q81SF6	Q81sf6 caenorhabdi
65	445	3.4	18534	5	Q81SF7	Q81sf7 caenorhabdi
66	440.5	3.3	1992	3	Q9P6T1	Q9p6t1 neurospora
67	438	3.3	2752	11	Q63461	Q63461 rattus norv
68	436	3.3	1151	13	Q57580	Q57580 gallus gall
69	434.5	3.3	9196	5	Q81Q87	Q81q87 drosophila
70	434	3.3	1783	4	Q15038	Q15038 homo sapien
71	434	3.3	10578	5	Q81SF5	Q81sf5 caenorhabdi
72	433.5	3.3	2084	3	Q9HEC9	Q9hec9 neurospora
73	429	3.2	1404	4	Q9BX49	Q9bx49 homo sapien
74	428	3.2	1404	4	Q92954	Q92954 homo sapien
75	428	3.2	2759	5	Q9VID9	Q9vid9 drosophila
76	424.5	3.2	1268	4	Q75046	Q75046 homo sapien
77	424	3.2	2089	4	Q86Z06	Q86z06 homo sapien
78	422	3.2	2090	4	Q96QC2	Q96qc2 homo sapien
79	421	3.2	2529	5	Q9VWC0	Q9vwc0 drosophila
80	419.5	3.2	2089	4	Q14676	Q14676 homo sapien
81	419	3.2	6994	5	Q17343	Q17343 caenorhabdi
82	419	3.2	6994	5	Q17490	Q17490 caenorhabdi
83	416.5	3.2	1791	4	Q60382	Q60382 homo sapien
84	416.5	3.2	3166	5	Q9W3Z0	Q9w3z0 drosophila
85	414	3.1	2137	4	Q15021	Q15021 homo sapien
86	412.5	3.1	1754	11	Q80U37	Q80u37 mus musculus
87	409.5	3.1	1422	6	Q95KU4	Q95ku4 canis famli
88	409	3.1	2376	5	Q966V1	Q966v1 drosophila
89	408.5	3.1	2376	5	Q9V5J0	Q9v5j0 drosophila

	90	405.5	3.1	1400	11	Q9ESU6	Q9ESU6 mus musculus
	91	405.5	3.1	1400	11	Q8VHF8	Q8VHF8 mus musculus
	92	405.5	3.1	1480	10	Q9LIE8	Q9LIE8 arabidopsis
	93	405.5	3.1	2382	4	Q9H4A3	Q9H4A3 homo sapien
	94	404	3.1	1274	5	Q20007	Q20007 caenorhabdi
	95	403	3.0	1110	13	Q91255	Q91255 petromyzon
	96	401	3.0	2656	5	Q9NE92	Q9NE92 leishmania
	97	400.5	3.0	1118	5	Q9VR13	Q9VR13 drosophila
	98	400.5	3.0	3084	12	Q8UZ11	Q8UZ11 pseudorabie
	99	396.5	3.0	1464	4	Q8N473	Q8N473 homo sapien
	100	396.5	3.0	2936	5	Q9NKP7	Q9NKP7 leishmania

ALIGNMENTS

RESULT 1

Q8QFM0	PRELIMINARY;	PRT; 1998 AA.
D Q8QFM0;		
C C Q8QFM0;		
I T 01-JUN-2002	(TREMELrel. 21, Created)	
T T 01-JUN-2002	(TREMELrel. 21, Last sequence update)	
I T 01-OCT-2003	(TREMELrel. 25, Last annotation update)	
E T SC:BZ7IM17.2.1	(Novel protein similar to mouse silencing mediator of retinoic acid and thyroid hormone receptor (SMRT)) (Fragment).	
E E SC:BZ7IM17.2.		
N N Brachydanio rerio (Zebrafish) (Danio rerio).		
S S Chordata; Metazoa; Craniata; Vertebrata; Euteleostomi;		
C C Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;		
C C Cyprinidae; Danio.		
X X NCBI_TaxID=7955;		
[1]		
N N SEQUENCE FROM N.A.		
P P Laird G.;		
A A Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.		
L L -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
-1- SIMILARITY:		
C C -1- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.		
R R EMBL; AL590153; CAD24407.1; -.		
R R GO; GO:0005634; C:nucleus; IEA.		
R R GO; GO:0003677; F:DNA binding; IEA.		
R R InterPro; IPR001005; Myb DNA binding.		
R R Pfam; PF00249; myb DNA-binding; 2.		
R R SMART; SM00717; SANT; 2.		
R R PROSITE; PSS0090; MYB_3; 1.		
R R Nuclear protein.		
R R NON TER		
TO NT		
SEQUENCE	1998 AA; 216897 MW; 4586606G3FCF9EE CRC64;	

[illegible]

1058	QY	-PPREVIKASPHADPDSAFS	YAPGCHPLPLGLDHTARVPLRPPTTISNPPPLISSAKHPS	1116				
615	DB	LSPRDMAKFS	---QEQLLHNP---SLASTLQPODRMAAVRPLHIPEPPLISSAK-PG	666				
1117	QY	VLEROIGAI	SOQMSVOLHVPY--SEHAKAPVGPVTWGLPLPMPDKKLA	PSPGVKQOLSP 1174				
667	DB	-----GSITQCTP	VQLHNP	SHGSEHGKVP--APGSLALSW-VQORXMGFPVVKQOLSP 718				
1175	QY	RCQAGPPRESL	GVPTAQEASVLRGTA	LGALSGVPGCSITTKGIPTRVPSDSAITR--GSITHGT 1233				
719	DB	RGASSSSQA	ENLSTQDS	TSRG--PMPAVQGSITKIGPTRVHQDPSPIVYRGSSITQGT 777				
1234	QY	PADVLKGTIT	RIIGEDSPSLDRG	EDSLPKGHVIVYEGKKGHVIVYEGGMSVTOCSKED 1293				
778	DB	PADVLKGTIT	RLINEDSPSR--ERSREDTP	SKGHVVYEGISGHILSFERRPN-----SKEE 832				
1294	QY	GRSSSGPPHET	AAPKRTYDMMEGRVGR--AISSAS	TEGLMGRAIPIPERHSPHILKQHH 1350				
833	DB	GCQGDVG----	AGMKRSYDMMEVG	ISSRVPIRDLSEGLIIRAMPDRDSPH-----PQL 883				
1351	QY	IRGSTIQIG	PRSYVEAQEDYLRE	EAKLKRCTGTPPPPSRDLTEAYKTQALGPLKUPA 1410				
884	DB	IRGSIQIG	IPR-----DDCQR	VEKMKREGSPSRGFGHS--TDALKSR-----S 927				
1411	QY	HGLVATVKE	AGRSIHETPREL	RHTPELPLAPRPLKEGSI	TOGTPLKYDTGASTTGSKK 1470			
928	DB	HESMVP	TVKEGGRS	IHILP-----PEGV	WIGKP-KGGSITQGTPLKQEPGCS---SKR 976			
1471	QY	HDVRSILIG	SPGRTFPVPH	PLVDMADARALERAC	YBES-LKSRPGTASSGSGSIARGAPVI 1529			
977	DB	HDVRSII	IASPR---PYSGL	PSHLDRPGSDRARI	EEAGCKRPSAVVAPSLSRVSPLA 1033			
1530	QY	V--PEIGK	PQOSPLTYEDH--CAP	PAGHLPRGSPVTWRE	PTPLQEGSLSSKASQDRK 1584			
1034	DB	LSGQSSRP	PHHSVGYEDHKRS	YPPSPH--RASPL	SGRENSQRAHEG--SKPQOQERK 1089			
1585	QY	LTSTPREI--	AKSPHSTVPSH	PHPTSPYEHLLRG	VGSDLYRSHIP-LAFDPTSIPRGI 1641			
1090	DB	ATPTPREMG	STKSP-LAVGE	TAN--LAFERIL	GL-GPDMVRSQMSHLTFFAAALPRGI 1144			
1642	QY	PLDAAA	YYLPRHLN	APNTVPHYLPPY	YLIRG-YPDTAAL	ENQCTIINDVITSQOMHNTA 1700		
1145	DB	PID--SAY	LPRHLAPG	GYPHYPYPLIRG	CGFETAAL	ENQTLFNDVITSQOMHNTA 1202		
1701	QY	TA--MAOR	ADMLRGLSP	RESSLAIN	AAGPRGIIDLSQV	PHLPVLVPTTGPATAMORL 1758		
1203	DB	AAAAAQR	PDLLRLGT	PREOSLN	LAYSPTPR-----GTS	ASPMORI 1243		
1759	QY	AYLP-TA	POFSSRHSS	SPSPGPGTHLTKPT	TTSSS-----ERE 1797			
1244	DB	TYIPGT	SPGPGAYNT	SPISPVGASHMS	KIQGSSAERERERERERERERERE 1303			
1798	QY	RDRDRDR	DREREK-----SIL	TSTTVTVEHAP	IWRPGT-EQSSGSSGSGGGGSS 1848			
1304	DB	RERDRDR	DRERKDRDR	DRDKLSHGN	VEHAPLWRFGA	PEQMS-----GSS 1353		
1849	QY	SRPASH	SHAHQHS	ISPTQDAL	QOORPSVLHNTGMKGIITAV	PEKPTVLRTSTSSPYR 1908		
1354	DB	SRPSPH	YSHQ-SPLSP	RPDQSLQOR	PSVLHNTSKSLSNS-BHNNP	SVLR-----1402		
1909	QY	PAATPP	PAT-----HCP	LGGLDGVYPTL	MEPVLVLLPKEAP	RVARPERPRADTGHAF 1960		
1403	DB	-----PPG	SARYQGFSGH	LQSGLPGEGY	PSGTDF-----S	AKDLR-----DGG---1442		
1961	QY	AKPPAR	SGL--EPAS	PSPK-----GSE	PRPLVPVSGHAT	IATPAKN---L 2002		
1443	DB	-KSHG	RGGLPKHQD	LSSSKSDS	KLASPCAG	GTVPSPY-----NOAP	VAHLPTSRGHPL 1496	
2003	QY	APHH-----	ASPDPP	APPASADP	PHREK	TKSPFSTQ	LELRSGLGHSYSGEV	VEVSP 2058
1497	DB	APESG	GGSSIOSRGD	SGSSASSREK	TQNKVMSIQ	EHLEALG-----KIT	WTAA	FINA 1552

Db 156 LYNQPSDTKVHENIKTNQVMRKULILFFKRNHARKQBQIKORYDQLMEAVEKKVDR 215
Qy 316 IENPRRAKESKVREYYEKOPFIRKQRELQRMQSRVQORGSLMSAARSEHEVSEI 375
Db 216 IENPRRAKESKVREYYEKOPFIRKQREQQERFQ-RVQGAGLSATARSBEHISEI 274
Qy 376 IDGLSEQENLEKQROLAVIIPMLYDQOORIKFIMNGLMADPMKYKYKDRQVMWSEQ 435
Db 275 IDGLSEQENLEKQROLAVIIPMLYDQOORIKFIMNGLMEDPMKYKYKDRQVMWTDH 334
Qy 436 EKETPREKPMOHKPNFGLIASFLERKTVACVLYLTKKNEYKSLVRRSV-RRRGKSQ 494
Db 335 EKEIFKDFIOPKPNFGLIASFLERKTVACVLYLTKKNEYKSLVRRSVYKGRGRNQ 394
Qy 495 QQQQQQQQQQQQQPPRSQEEKDEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 554
Db 395 QIARPSQEEKVEEKEE--DKAEKTEKKEEKKOBEKDEKESKENTKE---KOKIDGTA 449
Qy 555 GEDNDEKAVASKGRKTANSQRRKGRITRSMANENSEBALTPOOS-----AEL 604
Db 450 -ETEEREQATPRGRKTANSQRRKGRITRSMNTEAAAAASAAAAAATEEPPLPPPPPEP 508
Qy 605 ASMELNESSRWTEEREMETAKGLLEHGRNWSAIAARMVGSKTVSOCKNFYENYKKRONLDE 664
Db 509 ISTEPVTSRWTEEREMETAKGLLEHGRNWSAIAARMVGSKTVSOCKNFYENYKKRONLJDN 568
Qy 665 ILQOHLKWEKERNARRKKKAPAAASEEAAFPVVEDEEMEAASGVSEBEMVEEABAL 724
Db 569 LLQOHLKWEKERNARRKKKAPAAASEEAAFPVVEDEEMEAASGVSEBEMVEEABAL 619
Qy 725 HASGNEVPRGCSGPATVNNSSDTSIESIPSPH-TEAAKTONGPKPPATLGADGPPGPP 783
Db 620 -----AENSSDTSAPSPPVEAVK-----PSED 643
Qy 784 TPRRTSAPTEPTPASEATGAPPPAP-SPS-APPPVVPKEKEBETAAAPV---- 837
Db 644 SPENATSGNTEPAVELE-----PTTETAPSTSPSLAVSTKPADESSETVQNDOSISAET 699
Qy 838 -----BEGEKQPPAAELAVDTGKABEPPKSECTBEAEGPAGKDAEAAE 884
Db 700 AEQMDVDQOEHSAEBSVCDPPATK--ADSDVVEVRYPENHASKVEGDNTKERDLDR- 756
Qy 885 ATAEALKAKEKGGSGRATTAKSSGA-----PQ-DSDSATCSADEVDEAEGGDKNRLLS 939
Db 757 -----SEKVEPRDEDLVAQINAQRPESQSDNDSSATCSADE--DVDGEPRQRMF 806
Qy 940 P---RPSLLTPTGPRANAPQK--PLDLKOLKORAAAIPI----- 976
Db 807 PMSKPSLLNPTGSLV--SSPLKNPLDLPOLQHRAAVIPPVMSCTPCNIPITGVSVGYA 865
Qy 977 ---QVTKVHEPPREDAAPTKPAAPPQQNLOPESDAPQOQSSPRGKRSAPPADK 1032
Db 866 LYQRHAKMHESALLE-----EQRQOEIDLECRSSTSPGTSKSP-----NR 909
Qy 1033 EAFAAEAQKLPDGPCTWTSGLFPVPPPREVIKASHPADPSAFYAPGHPPLPLGLHDTA 1092
Db 910 E-----W-----EVLQAPAH-----QVITNLPGEVRLP-----TT 934
Qy 1093 RPVLPRPTIINPPPLISSAKHPSVLEROIGAISOG-----MSVLHVPYSEHAKAP-VG 1146
Db 935 RPTPPPLPIPSKTKTIVASEK-PGFI--MGSIISQGTGTLTSHNQASYTQETPKPSVG 991
Qy 1147 PVTMGLPLPMDPKKLAPPSSGVKQEQLSPRGQAGPPESLGVPTAOEASVLRGTALGSVPGG 1206
Db 992 SISGLPRQOESAKSATLPYIKQEFSPRSQNSQPEGLLV-RAQHEGVVRGTA-GAQEG 1049
Qy 1207 SITGIPSTRVPSDAITRGSITHGTPA-----DVLVYKGTITRIIGEDSPSLDRG 1258
Db 1050 SITRGTPTSISVESIPSLRGSITQGTALPOTGIPTEALVKGISRMPIEDSP--EKG 1107
Qy 1259 REDSLPKGHVIEYKKGHVLSYEGMSVTCQSKEDGRSSGPPHETAPKTYDWMGVRV 1318
Db 1108 REEAASGHGHVIEYKKGHVLSYEGMSVTCQSKEDGRSSGPPHETAPKTYDWMGVRV 1161

Qy 1319 GRAIS-----SASIEGLMGRAPRPHSPH-HLKEQHHRIGSITQGIIPRSVYEAQEDYL 1371
Db 1162 KQGMRESPVSAPLEGLICRALP--RGSPHSLDKERTVLSGSIQGTPTATTESFEDGL 1219
Qy 1372 BREAKLLKREGTTPPPPSRDLTEAYKQALGPLKPKAHEGLVATVKEAGRSIHEIPRE 1431
Db 1220 -KYPKQIKRES-----PPIRAFEGAI-----TKGP-YDG-ITTIKEMGRSIIHEIPRQ 1264
Qy 1432 EL-----RHTPELPLAPRLKEGSIQGTPLKYDTGASTTGSKGDVRSILGSPORTPP 1486
Db 1265 DILTQESKTEPVQSTRIPIEGSISQGTPIKFDNN-SQSAIKKXVKSILTGPSKLSRG 1323
Qy 1487 VHLDMVAD-ARALERACYE-----ESLSRPGTASSGSIARAGPVIVPELKGPROSP 1540
Db 1324 MPPELIVPENIKVVERGYEDVKAGETVRSRHTSVSSGSPVLST--LHEAPKAQLSP 1380
Qy 1541 LTYEDHGA-----PFAGHLPRGSPVTMREPTPRLOEGSLSSKA-SQPKLSTSTPRE--- 1591
Db 1381 GIYDDTSARRTPVSYQNTMSRGSPPMNR-----TSDVSNKSTNHERKSTLTPTORES 1433
Qy 1592 -IAKSPHSTVPEHHPHPIPSVYELLRGVSGVDLYRSHIPLAFDPTSIPRGIPLDAAAAYV 1650
Db 1434 IPAKSPVPGVDVUSH--SPFDPHHRGSTAGEVTRSHUPTHLP-AMFPHRALDPAAYVL 1490
Qy 1651 LPRHLAPNTYPHYLPYPIRGPDTAALEN-RQTIINDYITTSQQMHNATATAMAQRADM 1709
Db 1491 FORQLSTPGYPSQVLY-----AMENTROTILANDYITTSQQMNVN-----RPDV 1535
Qy 1710 LRGLSPRESSLALYAGPRGIIIDLSQVPHLPVLVPTPGTPATAMDRLAYLPTAPQSPS 1769
Db 1536 ARGLSPREQPLGLPYPA-TRGIIIDLTNMP--TILVPHPGGTSTPPMDRITVPGTITFP 1593
Qy 1770 SR-HSSPSPLSGPHTLTKPTTSSSERDRDRDREREKRSILSTTTTVEHAPIWR 1828
Db 1594 PRPNASMSPGHTHL---AAAASEREREREKERERER---IAASDOL-----YLR 1643
Qy 1829 PGTEQSSGSSSGSGSSSRPASHAHQHSIPSPRTQDA-LQORPSVLHNTMGKII 1887
Db 1644 PGSEQ-----PGRPGSHGVYRSPSP-SVRTQETMLQORPSVFGTNGTSVI 1688
Qy 1888 TAVPSPKTVLURSTSTSPV-----RPAATFPATHCPGLGTLGDVGYPTLMBPVLLPK--- 1940
Db 1689 TPLDPTAQLRIMPLPAGGPSISQGLPASRYNTAADA-LAALVDAASAAPQMDVSKTESK 1747
Qy 1941 -EAPRVARPERPRAD-----TGHAFLAKPPARSGLEPASS- 1974
Db 1748 HEARLEENLSRSAAVSEOOOLEQKLEVEKRSVQCLYTSAPPSGKP-----QPHSV 1802
Qy 1975 --PSKSGSEPRPLVPVSGHATIAITPAKN--LAPHASPDPPAPPASADPHREKTQSKP 2030
Db 1803 VYSEAGKDKGP--PPKSRYEELRTRKTTITTAANFIDVITRQIASDKDARERSQSD 1860
Qy 2031 PSIOELELRSLGYGSSYSPGVPSVPSLTHDKGLPKHLEELDKSHLEGLRPKQ 2090
Db 1861 SS-----SSLSSHRYETPSDAIEVISPASSPAPPOEKLOTYQPEVVKANQANDPTQY 1914
Qy 2091 PGVVKLGEAAHLPHLPLPESQSSPPLQTAG-----VKGHORVVTLAQHSIV 2142
Db 1915 EGP-----LHRYR---QOESPQQOLPSSQAEGMGQVPRTHRLITLADHICQI 1962
Qy 2143 ITQDYTRHH-----PQLSAPLAPLYSFPGA--SCPVLDIRRPPSLDLYLPPD-----H 2190
Db 1963 ITQDFARNQVSSQTPQO---PPTSTFNQSPALVSTPV---RTKTSNRYSPESQAQSVH 2016
Qy 2191 GAP-ARSGPH-----SEGGKRSPEPNKTSVLUGGEDGIEPVSPPEGMTPEGHRSVPL 2244
Db 2017 QRPGRSVSPENLVKRSRPGKSPERSHV---SSEPYEIPSPQ--VPVVEHKQDSLLL 2071
Qy 2245 LYRGEQTEPSRMGSKPGNTSQPAPFESKLTESNAMAIVKSKQEIKNKLNTHNREPEY 2304
Db 2072 LSQRBGAPAEQRNDARPGSISYLPSPFTKL-ENTSPMVKSKQEIFRKLJNSGGGSDM 2130

QY	2340	MGLEAIRKALMGKYDQWEESSPLSANAFNPLNASASLPAAMPITTAADGRSDHILTSPOG	2399
Db	429	MGLEAIRKALMGKYDQWEESSPLSANAFNPLNASASLPAAMPITTAADGRSDHILTSPOG	488
QY	2400	GGKAKVSGRPSRRKAKSPAPGLASGDRPPSVSVHSEGCNRRTPLTNRWEDRPPSSAGS	2459
Db	489	GGKAKVSGRPSRRKAKSPAPGLASGDRPPSVSVHSEGCNRRTPLTNRWEDRPPSSAGS	548
QY	2460	TPFPYNPLIMRLQAGVNASPPPPGCLPAGSGPLAGPHAWDEEPEKPLLCQYETLSDSE	2517
Db	549	TPFPYNPLIMRLQAGVNASPPPPGCLPAGSGPLAGPHAWDEEPEKPLLCQYETLSDSE	606
RESULT 6			
ID	Q80ZV7	PRELIMINARY; PRT; 631 AA.	
AC	Q80ZV7	01-JUN-2003 (TremBLrel. 24, Created)	
DT	01-JUN-2003	(TremBLrel. 24, Last sequence update)	
DT	01-OCT-2003	(TremBLrel. 25, Last annotation update)	
DE	Similar to nuclear receptor co-repressor 2 (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Eye;		
RA	Strausberg R.;		
RL	Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC047524; AAH47524.1; -		
DR	GO; GO:0004872; F:receptor activity; IEA.		
KW	Receptor.		
FT	NON_TER		
SQ	SEQUENCE 631 AA; 66869 MW; CC1F52630A984D6E CRC64;		
Query Match 22.8%; Score 3012; DB 11; Length 631;			
Best Local Similarity 90.6%; Pred. No. 2.2e-137;			
Matches 576; Conservative 11; Mismatches 39; Indels 10; Gaps 6;			
QY	1887	ITAVEPSKPTVL---RSTSTSPVRPAATFPFATHCPGLGTLGVYPTLMEPVLPPKEAP	1943
Db	1	VTSVEPTGTVLRWARSTSTSPVRPAATFPFATHCPGLGTLGVYPTLMEPVLPPKETS	60
QY	1944	RVARPERPADTGHAFLAPKPARSGLEPASSKSEPRPLVPVSGHATARTPAKNLA	2003
Db	61	RVARPERPRVDAGHAFLTKPPAR---EPASSPSKSEPRSLAPPSSHTARTPAKNLA	117
QY	2004	PHASDPDPAPPASADPHREKTQSPFSGIQLERLSLYH-GSSYSPEGVEPVSPVSSP	2062
Db	118	PHASDPDPA-PTASADLHREKTQSPFSGIQLERLSLYHSGAGYSPDGVPISPVSSP	176
QY	2063	SLTHDKGLPKHLEBKSHLEGELRPQGPVKLGEEAAHLPHLRPLPESQSSPLLOT	2122
Db	177	SLTHDKGLSKPLEBEKSHLEGELRHQGPCKMLSAEAAHLPHLRPLPESQSSPLLOT	236
QY	2123	ARGVKHORVVVTLAHLSEVITQDYTRHHPQOLSAPLPAPLYSFGASCFLVLDLRPPSD	2182
Db	237	APGLKHQRVVVTLAHLSEVITQDYTRHHPQOLSGPLPAPLYSFGASCFLVLDLRPPSD	296
QY	2183	LYLPPPDHGAARGSPHSEGGKSPFNKTSVLGGEDGIEPVSPPEGTEPHSRSAY	2242
Db	297	LYLPPPDHGTARGSPHSEGGKSPFNKTSVLGGSEDAIEPVSPPEGTEPHARSTAY	356
QY	2243	PLLYRDEQTEPSRMGSKSPGNTSOPAPFSKLTESNVMVSKQOEINKLNTNRNEP	2302
Db	357	PLLYRDEQGEPE-RMGSKSPGNTSQPPAFFSKLTESNVMVSKQOEINKLNTNRNEP	415
QY	2303	EYNISQPGTEIFNNPAITGTGLMTYRQAVQEHASTNMGLEAIRKALMGKYDQWEESSP	2362
Db	416	EYNIGQPGTEIFNNPAITGAGTLCRSQAVQEHASTNMGLEAIRKALMGKYDQWEESSP	475
QY	2363	LSANAFNPLNASLPAAMPITTAADGRSDHILTSPOGGGKAKVSGRPSRRKAKSPAPGL	2421

Db	476	LGANAFNPLNASLPAAMPITTAADGRSDHILTSPOGGGKAKVSGRPSRRKAKSPAPGL	535
QY	2422	ASGDRPPSVSVHSEGCNRRTPLTNRWEDRPPSSAGSTFPFPYNPLIMRLQAGVNASPPPP	2481
Db	536	ASGDRPPSVSVHSEGCNRRTPLTNRWEDRPPSSAGSTFPFPYNPLIMRLQAGVNASPPPP	595
QY	2482	PGLPAGSGPLAGPHAWDEEPEKPLLCQYETLSDSE	2517
Db	596	PGLAAGSGPLAGPHAWDEEPEKPLLCQYETLSDSE	631
RESULT 7			
ID	Q86YY0	PRELIMINARY; PRT; 914 AA.	
AC	Q86YY0	01-JUN-2003 (TremBLrel. 24, Created)	
DT	01-JUN-2003	(TremBLrel. 24, Last sequence update)	
DT	01-OCT-2003	(TremBLrel. 25, Last annotation update)	
DE	NCOR isoform b.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yu L.;		
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF303586; AAC32942.1; -		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003677; F:DNA binding; IEA.		
DR	InterPro; IPR001005; Myb DNA binding.		
DR	Pfam; PF00249; myb DNA-binding; 2.		
DR	SMART; SM00717; SANT; 2.		
DR	PROSITE; PS00090; MYB 3; 1.		
SQ	SEQUENCE 914 AA; 103812 MW; 619057D2BD26480C CRC64;		
Query Match 16.5%; Score 2184; DB 4; Length 914;			
Best Local Similarity 51.4%; Pred. No. 3.1e-97;			
Matches 490; Conservative 110; Mismatches 208; Indels 146; Gaps 26;			
QY	137	KDRSLTKLE-PVSPSPPHDPELPLPPLPSKEELIQNMDRVDREITWVEQOISKLKK	195
Db	36	QDPAFGKGHEAPSPISGQCGDDQNASPSKLSKEELIQMDRVDREIAKVEQOILKLLK	95
QY	196	KQOLEEAAKPEPEKPVSPPIESKHSRLVQIYDENRKKAAEAHRLLEGLGPOVELP	255
Db	96	KQOLEEAAKPEPEKPVSPPVVEQKHSIVQIYDENRKKAAEAHRLLEGLGPKVELP	155
QY	256	LYNOPSDTROVHENIKINOAMRKKLILYFKRRNHARKQKQFCORYDQLMLEAKKVER	315
Db	156	LYNOPSDTKVHENIKTNQVNRKKLILYFKRRNHARKQREQKICORYDQLMLEAKKVER	215
QY	316	IENPRRAKESKVREYVEKQFPEIRKQELQERMOSVQRCGSLSMSAARSEHSEI	375
Db	216	IENPRRAKESKTRVEYVEKQFPEIRKQEQERFQ-RVQRCGAGLSATTARSEHSEI	274
QY	376	IDGLSEQENLEKONROLAVIPMLYDADQORIKFINNGLMADPMKYVKDQVNMWSEQ	435
Db	275	IDGLSEQESNKQWRQUSVIPMMFDABQRARVFINNGLMEDPMKYVKDQVNMWTDH	334
QY	436	EKETFRKFMQHPKNFGLIASFLERKTVAECVLYYLTKKNNYKSLVRSY-RRRKSQ	494
Db	335	EKEIFKDKFIQHPKNFGLIASYLERKSVPCVLYYLTKKNNYKALVRRNYGKRRNQ	394
QY	495	QQQQQQQQQQQQQQQPMPSQEKDEKEKEKEKEKEPEVENDKDLLEKTDTS	554
Db	395	QIARPSQEKVEKEE--DKAEKTEKKEEKKDEKEDSKENTKE---KDKIDGTA	449
QY	555	GEDNDEKAVASKGRKTANSQRRKGRITRSWANEANSEEAITPQOS-----AEL	604
Db	450	-EETEERQATPRGRKTANSQRRKGRITRSMTNEAASAAAAATEEPPPLPPPEP	508

Qy	2439	CNRRTPLNRVWEDRPSAGSTPPYPNPLIMRLQAGVMASPPPLGAPAGSPL--AGPH-	2495
Db	1646	YHRQTP--GWAWEDRPSSTGTFQFYNPLTIRM-----LSSTPTQIACAPSAITQAAPHQ	1699
Qy	2496	--HAWDEERKPLLCSQYETLSDSE	2517
Db	1700	QNRWIEREPAPLLSAQYETLSDSD	1723
RESULT 10			
ID	Q8VDE8	PRELIMINARY; PRT; 1399 AA.	
AC	Q8VDE8;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	RIP-13.		
GN	NCOR1 OR RIP13.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RC	MEDLINE=97120602;		
RA	Seol W., Mahon M.J., Lee Y.K., Moore D.D.;		
RT	"Two receptor interacting domains in the nuclear hormone receptor		
RT	corepressor RIP13/N-COR.;"		
RL	Mol. Endocrinol. 10:1646-1655(1996).		
DR	EMBL; L78294; AAL40135.1; -.		
DR	MGD; MGI:1349717; Ncor1.		
SQ	SEQUENCE 1399 AA; 151393 MW; D42406CA592BEE20 CRC64;		
Query Match 13.9%; Score 1839; DB 11; Length 1399;			
Best Local Similarity 35.5%; Pred. No. 2.3e-80;			
Matches 557; Conservative 210; Mismatches 517; Indels 284; Gaps 72;			
Qy	1057	VPPREVIKASPHADPSAFSAPPCHPLPLGLHDTARVLPRLPPTISNPPPLISSAKHPS	1116
Db	8	VPP--VLQAPH-----QVITNLPEGVLP-----TTRTRPPPPPLIPSSKTTVASEK-PS	55
Qy	1117	VLERQIGAIQSG-----MSVLHVYPYSEHAKAPGVFTVMGLPMDPKKLAPFSGVKQEQ	1171
Db	56	FI--MGGSISQCTPTCYLSSHNOAYPOEAPKPSVGSISLGLPRQQUESTKAAPLTIKQEE	113
Qy	1172	LSPRQAGPPSELGVPTAQEASVLGRTALGSPVGGSIKGPSTRPVSDSAITYRGSITH	1231
Db	114	FSPRSQSQPEGLLV-RAQHEGVVRGTA-GAVQEGSITRGTPASKISVETISSLRGSITQ	171
Qy	1232	GTPA-----DVLTKGTITRI-IGEDSPSLDRGREDSLPKGHVYIEGKKGHVLSEY	1282
Db	172	GTPALPQAGIPTREALVGPVSRMPLEESSPEKV-----REEAASKGHVYIEGSGHLLSY--	226
Qy	1283	GMSVTQCSKEDGRSSSGPPHETAAKRTYDMGVRGVAISSASIEGLMGRAIPPERHSP	1342
Db	227	-----DSLICRALP--RGSP	239
Qy	1343	H-HLKEQHHRGSIQGIIPRSVVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTAQ	1401
Db	240	HSDLKERTVLSGSIHQGTFRATAESFEDGL-KYPKQIKRES-----PPIAFEGAI----	289
Qy	1402	LGPLKLKPAHGLVATVKEAGRSIHEIPREL-----RHTPELPLAPPLKEGSTTQCTP	1456
Db	290	---TKGKP-YDG-ITTIKEMGRSHEIPRODILLTQESRKTPEVVQSTRPIIEGSIISQGT	344
Qy	1457	LKYDTGASTGSKKHVRSLSIGSPGRTFPVPHPLDVWAD-ARALERACYE-----ESLKS	1510
Db	345	IKFDNN-SGQSAIKHNKSLITGFSKL--PRGMLEIVPENIKVVERGYEDVKAGEPVRA	401
Qy	1511	PGTASSSGGSIAGAPVIVPELKGPROSPLTYEDHGA-----PFAGHLPRGSPVTMREP	1565

Db	402	RHTSVVSGSPVLST---LHEAPKAQLSPGLYDDSSARRTPTVSVYONTISRSGSPMNR--	456
QY	1566	TPRLQEGSLSSK-ASQDRKLTSTPRE---IAKSPHSTVPEHHPHPISPYEHLLRGVSG	1620
Db	457	-----TSDVSSKSASHERKSTLTPTQRESIPAKSPVPCVDPVVSH--SPFPDHRSSAA	509
QY	1621	VDLYRSHPLAPDPTSIPRGIPLD-AAAYYLPRHLAPNTYPPHYLIRGYPDTAAL	1679
Db	510	GEVYRSHLPTLDP-AMPFHRALDPAAAYLLQRLSPTPGPSPQYLY-----AM	559
QY	1680	EN-RQTIINDYITSOQMHNTATAMAQRADMLRGLSPRESSIALNVAACPRGIIDLQVP	1738
Db	560	ENTRQTIINDYITSOQMOVNL-----RPDVTIRGLSPREQLGLPYPA-TRGIILDLTNMP	612
QY	1739	HLPLVLPPTPGTATAMDRLAYLPTAPQPFSSR-HSSGSLSPGGTHLTKPTTTSSSRE	1797
Db	613	P-TILVPHAGGTSTPPMDRIYIYFGQVTFPPRPYNAASLSFGCHPTHL---AAAASAERE	668
QY	1798	RDRDRDRDREREKSILTSTTTTVEHAP---IWREGTQSGSGSSGGSGSSSRPASH	1854
Db	669	REREREKER	714
QY	1855	SHAHQHSIPSPRTQDA-LQORPSVLHNTGMKGIIITAVEPSKPTVLRSTSTSPV----	1909
Db	715	GYVRSPSP-SVRTQETILQORPSVFGTNGTSVITPLDPTAQLRIMPSPSGGFSISQGLP	773
QY	1910	AATPPATHCPGLGGLDGVYPTLMBPVLLPKEAPR-----VARPERPRADTG	1956
Db	774	ASRYNTAADA-LAALVDAASAPQMDVSKTKESKEAARLENLRSAANVSEQQOLEQ	832
QY	1957	HAFIAKPPARSGLBPASSPSKSGEPRLV-----PPVSGHATITARTPAKN--LA	2003
Db	833	NLEVEKRSVCQVCTSSALPSGKAQPHASVYSEAGKDKGPPPKRYEELRTRGTTITA	892
QY	2004	PHASDPDPAPASADPHREKTQSKPFSIQEELRSLGYHCSSVSPGCVSPVSPSSPS	2063
Db	893	ANFIDVIITRIQIAQDKARERGSSQSSDSS-----SSLSSHRYETASDAIEVSPASSA	946
QY	2064	LTHDKGLPKHLELDKSHLEGELRPKQPGVKLGGEAAHLPHLR-----PLPESOPSSSP	2118
Db	947	PPQEKQAYQDMVKANOAEENSTROYEGP-----LHHYRSQQSPSPSQOQPLPP	997
QY	2119	LLQT--APGVKGHORVWTLAQHISEVITQDYTRHHPOQLSAPLAPLYSFGQA--SCPVL	2174
Db	998	SSQSEGMCQVPRTHRLITLADHIQCIITQDFARN--QVPSQPTSTFTQTSALSSTPV-	1054
QY	2175	DLRPPSDLYLPPDP-----HGAPA-RGSPHS-----EGKRSPEPNKTSVLGGGEGDIEP	2224
Db	1055	--RTKTSRYSPEQSQTVLHPRPGRPVSPENLVKSRGPFKSPERSHI---PSEPYEP	1109
QY	2225	VSPEGMTEPG-HSRSAVYPLLRYDGEQTEPSRMGSKSPGNTSQPAPFSSKLTESNAMY	2283
Db	1110	ISPPQG---PAVHEKQDSMLLSQGVDPAPQRSDSRSPGSIYLPSPFTKL-ESTSPMV	1165
QY	2284	KSKQKQINKKLNTNRNEPEYNIISQGTETIFNMPAITGTLMTYRSQVQEHASTNMGLE	2343
Db	1166	KSKQKQETFRKLNSSGGSDMAAQAQGTETIFNLPAVTTSGAVSSRSHSFADPAS-NLGLE	1224
QY	2344	AIIRKALMGKYDQWEE-----SPLSANAFNPLNASLASLPAAMPDITADGRSDHTLSP	2397
Db	1225	DIIRKALMGSPDDKVEDHGVVMSPV---GIMPGSASTSV-----VTSSEARRDGEPS	1276
QY	2398	GGG-GKAKVSGRPSRRKAKSPAPGLA--SGDRPPSVSVHSEDCNRRITPLTNRWEDRP	2454
Db	1277	HAGVCKPKLINKSNRKSPKIPQOSYLGTERPSVSVHSEGDVHRTQP--GWAWEDRP	1334
QY	2455	SSAGSTPPPNPLIMRLQAGVMASPPPGLPAGSGPL--AGPH---HAWDEEPKPLCSQ	2509
Db	1335	SSTGSTQFPYNPLTIRM---LSSTPTQIACAPSAITQAAPHQNRWIREPAPLLSAQ	1390
QY	2510	YETLSDSE	2517
Db	1391	YETLSDSD	1398

RESULT 11

Q86W52	PRELIMINARY;	PRT;	550 AA.
ID	Q86W52		
AC	Q86W52		
DT	01-JUN-2003 (TREMBlrel. 24, Created)		
DT	01-JUN-2003 (TREMBlrel. 24, Last sequence update)		
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)		
DE	Similar to nuclear receptor co-repressor 1 (Fragment).		
DE	Homo sapiens (Human)		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID=9606;			
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=lung;		
RA	Strausberg R.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBDJ databases.		
DR	EMBL; BC050594; AAH50594.1; -		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003677; F:DNA binding; IEA.		
DR	GO; GO:0004872; F:receptor activity; IEA.		
DR	InterPro; IPR001005; Myb DNA binding.		
DR	Pfam; PF00249; myb DNA-binding; 1.		
DR	SMART; SM00717; SANT; 1.		
DR	Receptor.		
KW	NON TER		
FT	SEQUENCE 550 AA; 550		
FT	SEQUENCE 550 AA; 64673 MW; 362A54A80A5AE286 CRC64;		
QY	Query Match	12.8%; Score 1687; DB 4; Length 550;	
QY	Best Local Similarity	62.5%; Pred. No. 1.5e-73;	
QY	Matches	356; Conservative 59; Mismatches 103; Indels 52; Gaps 12;	
QY	16 EPRYPHSLSYVQIARTHTDVCGLLEYQ--HHSRDYASHLSGSIQQRERRPSLSEFQ 73		
QY	17 QSRYPHSHVQYTPFTRHQEFAPVDYRSHLEVSAQLQLQQQQQLRRRPSLSEFH 76		
QY	74 PGNERSQLHLRSHSLYLPGLCKSMETIESKRRLLELLPD-----PILRPS 122		
QY	77 PGSDRPOE--RRTSYEPFHGSPVDHDSLESKRPLEQVSDSHFQVSAALPLVHPLP 134		
QY	123 LLATGQPGAGSEDTKORSLSLTKGLE-PVSPSPPHDTPDELIVPRLSKEELIIONMDVR 181		
QY	135 ---EGLRA-SADAKKDPAGFGKEHAFSSPISGQCGDDQNGSPKLSKEELIQSDVR 190		
QY	182 EITWVEQOISKLKKQOOLEEAAKPEPEKPVSPPIESKHSIVQIYIDENRKAEEA 241		
QY	191 EIAKVEQOILKLLKKQOOLEEAAKPEPEKPVSPPVQKHSIVQIYIDENRKAEEA 250		
QY	242 HRILEGLGPQVELPLYNQPSDTRQYHNTIKINQAMKKLILFKRNHARKQWKFQ 301		
QY	251 HKIFEGGLPKVELPLYNQPSDTRQYHNTIKINQAMKKLILFKRNHARKQWKFQ 310		
QY	302 YDQLEALEKKVEIENNNRRKESKRVVEYKQPEIRKQRELOERQWQVQSGSL 361		
QY	311 YDQLEAWEKKVDRIENNNRRKESKRVVEYKQPEIRKQRELOERQWQVQSGSL 369		
QY	362 SMSAARSHEVSEIIDGLSEQENNEKQMRQLAVIPMLYDADQRIKFIMNGLMADPMK 421		
QY	370 SATIARSEHSEIIDGLSEQENNEKQMRQLAVIPMLYDADQRIKFIMNGLMADPMK 429		
QY	422 VYKQVQNMWSEKETFREKFMQHPKNGLTASFLERKTVAECVLYYLTKCNENYKS 481		
QY	430 VYKQVQNMWVTDHEKEIFDKFIQHPKNGLTASFLERKTVAECVLYYLTKCNENYKA 489		
QY	482 LVRRSY-RRRGKSQQQQQQQQQQQQQQQQMPRSSQEEKDEKEKEKEKEKEKEKE 540		
QY	490 LVRRSY-RRRGKSQQQQQQQQQQQQQQQQMPRSSQEEKDEKEKEKEKEKEKEKE 531		
QY	541 DKEDLLKKTDTSDGNDKEAVASKRK 570		
QY	532 -----EKKDEK--DSKKK 549		

RESULT 12

Q9NSZ0	PRELIMINARY;	PRT;	1047 AA.
ID	Q9NSZ0		
AC	Q9NSZ0		
DT	01-OCT-2000 (TREMBlrel. 15, Created)		
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-JUN-2003 (TREMBlrel. 24, Last annotation update)		
DE	Hypothetical protein (Fragment).		
DE	KFZP434M075.		
GN	Homo sapiens (Human)		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID=9606;			
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=restis;		
RA	Duesterhoef A.; Lauber J.; Mewes H.W.; Gassenhuber J.; Wiemann S.;		
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBDJ databases.		
DR	EMBL; AL137641; CAB70854.1; -		
DR	PIR; T46489; T46489.		
KW	Hypothetical protein.		
FT	NON TER		
FT	SEQUENCE 1047 AA; 113817 MW; B5A1EDE938B7222A CRC64;		
QY	Query Match	9.6%; Score 1272; DB 4; Length 1047;	
QY	Best Local Similarity	35.0%; Pred. No. 3.6e-53;	
QY	Matches	401; Conservative 149; Mismatches 393; Indels 202; Gaps 54;	
QY	1470 KHDVRLIGSPGRTFPVPHLDV		


```
Qy 2073 HLELDKSHLGEIRPKQGVKVLGGEAAHPLHLPLPESQPSPLLOTAPG----- 2125
Db : : : : :
Qy 600 QPEVVKANQANEDPTROVEGP-----LHHYRP-----QOESPQOQLPSSQAEGMG 647
Db : : : : :
Qy 2126 -VKGHQVVTTLAQHISEVITDYTRHH-----POOLSAPLAPLYSPGPA--SCPVLDR 2177
Db : : : : :
Qy 648 QVPRTHRLITLADHICQIITQDFARNQVSSQTPQO---PPTSTFQNSPALVSTPV---R 701
Db : : : : :
Qy 2178 RPPSDLVLPPDP-----HGAP--ARGSPH-----SEGGRSPPEPNKTSVLGGEGDIEPVS 2226
Db : : : : :
Qy 702 TKTSNRYSPESQAQSVHHORPGSRVSPENLVKSRGSRPGKSPERSHV---SSPEYEPIS 758
Db : : : : :
Qy 2227 PPEGMTPECHRSRVAVYLLYRDGQTEPSRMGSKSPGNTSQPPAFPSKLTESNAMYKSK 2286
Db : : : : :
Qy 759 PPQ--VPVVEHEKQDLSLLSQGAEPAEQRNDARSQGISYLPSPFTTKL-ENTSPMVKSK 815
Db : : : : :
Qy 2287 KOEINKKLTNHNREPEYNIQPGTEIENMFAITGTGLMTYRSQAVQEHASTNGLAII 2346
Db : : : : :
Qy 816 KOEIFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSGSVSRGHSFADPAS--NLGLEDI 874
Db : : : : :
Qy 2347 RKALMGKYDQWEE-----SPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGG 2400
Db : : : : :
Qy 875 RKALMGSDDKVEDHGVMSQPMGV---VPGTANTSV-----VTSGETRREGDPSPHSG 926
Db : : : : :
Qy 2401 G--KAKVSGRPSRKAKSPAP--GLASGDRPPSVSVHSEGDGNCNRRTPLTNRVWEDRPSS 2456
Db : : : : :
Qy 927 GVCKPKLISKNSRKSPIPGQVGLGTERPSSVSVHSEGDYHRQTP--GWAWEDRPSS 984
Db : : : : :
Qy 2457 AGSTFPFYNPLMRLOAGWASPPPGPLPAGSG--PLAGPH---HAWDEEPKPLICSQVET 2512
Db : : : : :
Qy 985 TGSQFPYFNPLTRM---LSSTPTPTIACAPSAVNOAAPHQONRIWEREPAPLLSAQVET 1041
Db : : : : :
Qy 2513 LSDSE 2517
Db : : : : :
Qy 1042 LSDSD 1046
Db : : : : :

RESULT 13
Q86YV1 PRELIMINARY; PRT; 959 AA.
AC Q86YV1
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE NCOR isoform c.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Yu L.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF303585; AAC32941.1; -.
SQ SEQUENCE 959 AA; 104500 MW; 35F97ED97CFB1CDD CRC64;

Query Match 8.9%; Score 1172.5; DB 4; Length 959;
Best Local Similarity 35.6%; Pred. No. 2e-48;
Matches 366; Conservative 127; Mismatches 355; Indels 181; Gaps 48;

Qy 1573 SLSSSKA-SODRKLSTPRE----IAKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSH 1627
Db : : : : :
Qy 27 TISSNKSTNHERKSTLTPTQRESIPAKSPVGVDPVVSH--SPDPHHRGSTAGEVYRSH 84
Db : : : : :
Qy 1628 IPLAFDPTSPRGIPLDAAAAYILPRHIANPTPHLYPPVILIRGYDPTAALN-RTIIL 1686
Db : : : : :
Qy 85 LPHLDP-AMPFHRALDPAAYLFQRLSPPTPGPESQVQLY-----AMENTROTIL 134
Db : : : : :
Qy 1687 NDYITSOQMHNTATAMAQADMLRGLSPRESSALANVAGPRGIIDLQVPHLPVLVPP 1746
Db : : : : :
Qy 135 NDYITSOQMVNL-----RPDVARGLSPREQPLGLPYPA--TRGIIDLTMNPP--TILVPH 186
Db : : : : :
```

```
Qy 1747 TPGTAPATAMDLAYLPTAQPOFSSR--HSSSPLSPGPGPHLTKPTTTSSSERERDRDRD 1805
Db : : : : :
Qy 187 PGGSTPMDRITITPGTQITFPPRPYNASMSGPHPHL---AAAAGAEREREREKE 243
Db : : : : :
Qy 1806 RDREREKSLTSTTTVEHAPIWRBCTEQSSGSSGSGGSSRPPASHAHQHSPISP 1865
Db : : : : :
Qy 244 RERER---IAAASDL-----YLRPGSEQ-----PGRPGSHGVRSPP--SV 281
Db : : : : :
Qy 1866 RTQDA--LQORPSVLHNTGMKGIITAVBPSKPTVLTRSTSSPV---RPAATFFPPATHCP 1920
Db : : : : :
Qy 282 RTQETMLQORPSVFGINGTSVITPLDPTAQLRMPLPAGGPSISQGLPASRYNTAADA- 340
Db : : : : :
Qy 1921 LGTGLDGYVPTLMBPVLLPK-----EAPVARPERPRAD----- 1954
Db : : : : :
Qy 341 LAALVDAASAPQMDVSKTKESKHEAARLEENLRSAVSEQQOQLEQKLTVEKRSVQC 400
Db : : : : :
Qy 1955 --TGHAFKAPKPARSGLEPASS---PSKGSPPRLVPVSGHATARTPAKN--LAPPHA 2007
Db : : : : :
Qy 401 LYTSSAFPSGKP-----OPHSSVVYSEAGDKGP--PPKSYEBELRTRGKTITTAANFI 453
Db : : : : :
Qy 2008 SPDPAPPASADPHREKTKOSKPFISIQEELRLSLGYHGSSYPGEVPEVSPVPSLSLTHD 2067
Db : : : : :
Qy 454 DVIITROIADSKDAREGSSQSDSS-----SSLSSHRYETPSDAIEVISPASPPAQE 507
Db : : : : :
Qy 2068 KGLPKHLELDKSHLEGELRPQGVKVLGGEAAHPLHLPLPESQPSPLLOTAPG-- 2125
Db : : : : :
Qy 508 KLQTYQPEVVKANQANEDPTROVEGP-----LHHYRP---QOESPQOQLPSSSQ 555
Db : : : : :
Qy 2126 -----VKGHQVVTTLAQHISEVITDYTRHH-----POOLSAPLAPLYSPGPA--SCP 2172
Db : : : : :
Qy 556 ABGMGQVPRTHRLITLADHICQIITQDFARNQVSSQTPQO---PPTSTFQNSPALVSTP 612
Db : : : : :
Qy 2173 VLDLRRPSPDLYLPPDP-----HGAP--ARGSPH-----SEGGRSPPEPNKTSVLGGEGD 2221
Db : : : : :
Qy 613 V---RTKTSNRYSPESQAQSVHHORPGSRVSPENLVKSRGSRPGKSPERSHV---SSEP 666
Db : : : : :
Qy 2222 IEPVSPPEGMTEPGHRSRVAVYLLYRDGQTEPSRMGSKSPGNTSQPPAFPSKLTESNSA 2281
Db : : : : :
Qy 667 YEPISPPQ--VPVVEHEKQDLSLLSQGAEPAEQRNDARSQGISYLPSPFTTKL-ENTSP 723
Db : : : : :
Qy 2282 MVKSKQOIKKLTNHNREPEYNIQPGTEIENMFAITGTGLMTYRSQAVQEHASTNMG 2341
Db : : : : :
Qy 724 MVKSKQOEIFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSGSVSRGHSFADPAS--NLG 782
Db : : : : :
Qy 2342 LEAIIKALMGKYDQWEE-----SPPLSANAFNPLNASASLPAAMPITAADGRSDHTLT 2395
Db : : : : :
Qy 783 LEDIIRKALMGSDDKVEDHGVMSQPMGV---VPGTANTSV-----VTSGETRREGDP 834
Db : : : : :
Qy 2396 SPGGG-QKAKVSGRPSRKAKSPAP--GLASGDRPPSVSVHSEGDGNCNRRTPLTNRVWED 2452
Db : : : : :
Qy 835 SPHSGVCKPKLISKNSRKSPIPGQVGLGTERPSSVSVHSEGDYHRQTP--GWAWED 892
Db : : : : :
Qy 2453 RPSAGSTPPFYNPLMRLOAGWASPPPGPLPAGSG--PLAGPH---HAWDEEPKPLICS 2508
Db : : : : :
Qy 893 RPSSTGSTQFPYFNPLTRM---LSSTPTPTIACAPSAVNOAAPHQONRIWEREPAPLLSA 949
Db : : : : :
Qy 2509 QYETLSDSE 2517
Db : : : : :
Qy 950 QYETLSDSD 958
Db : : : : :

RESULT 14
Q9VYK0 PRELIMINARY; PRT; 3604 AA.
AC Q9VYK0
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG4013 protein.
GN SMR OR CG4013.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```


Db 1252 PPEISAGEAKSKNAEBEAAATAGAAVATVATAGTATGASAAASAGAEATTATGATATAAKG 1311
Qy 603 ---ELASMELNESRWTTEE---METAKKLELHGRNWSAIAIMVSGKTVSOCKN 651
Db 1312 VGPETATEPAGTAAGADSPDANDPLAKTASAINAEGN-----AIGNSSSSSSN 1365
Qy 652 ---FYFNKKQKQNLDE-----ILOQHK 670
Db 1366 ATGASAPVQVTLNGFKPGYQTVVMANVKASTGGDDSGANAGGAAPGSLAATNASIATSG 1425
Qy 671 LKMKERNAREKKKAPAAAEAEAFPPVDEEMEGSV-----SGN----- 713
Db 1426 DKIVKTPSSR-----APNSTSSTAA-----NEGSSGAGVNTYGHATTAGTAYLGOKLKA 1475
Qy 714 -BEEMVEAEALHASGNEVPR-----GECGSPAT--VNNSDSTESIPSPHTBAAKDTG 763
Db 1476 AQVEGLGAGNELHSDVSESKKRFLNSGEAGGNATSAMTNSSTSGSMNINSHGLKANA 1535
Qy 764 QNG-----PKPPAT----- 772
Db 1536 KDGSMMAKTSMASTSSASVVVSTPSSASSLSASSMALLISAASVMSTAAGATSSSTAT 1595
Qy 773 -----LGADGPPG-----PPTPPRTSRA-- 792
Db 1596 TTATASALSPLADGSGNSVMNANEILADGKOKLASCFCVKACACPRTRPLKKGRGQQ 1655
Qy 793 ---PIEPTPAS-----EATCAPTPPPAPPSPAPPV-----VPKEEKEETAAA 834
Db 1656 YGIDEITIPAGARVCNSQCKSVSRYPNCPPLTCPNPKDRAQLRNIPSLFE-----LA 1711
Qy 835 PVVEE---GEQKPPAA-----BELAVDTGKA-----BEPVKSE 865
Db 1712 PEVRDPLMAEQIIPPHATRCSSACLMRIRRLKLDPLQNLTDGSSGAGSGGDETDVSTS 1771
Qy 866 CTEEAEGCPAKGDAEAAEATAEGALKAEK----- 895
Db 1772 SCDREPG---GSDTASVESPEN---LQHKLSLTMVKQKQKQKQKQKQKQKQKQKQKQ 1826
Qy 896 -----KEGSGRAVTA----- 906
Db 1827 SQOPPPAPQOQKSGSGRGDQGTPLIITPRMSSKSGSGGAQTAGDNERLLPPAAGQA 1886
Qy 907 -KSGAPQDSDSATCSADEVDE-----AEGDKNRLSLPRP---SL 944
Db 1887 PKQKTSBEEVDSSATETADEENENSPANRQSPKVLFGHGHGHGHGHANNVAGLPVAVNM 1946
Qy 945 LTPGDPANASPOK---PLDLKQLKQRAAIPPIQ--VTKVHEPPREDAAPTKEAPAP 999
Db 1947 GTGGGVQFGGAAGQVNGPISM-----RREAVNVQDCVFSVIERSLKHKGK----- 1993
Qy 1000 PPQNLQESDAPQOPG-----SSPRGKSRSPAPPADKEAFAAEAOKLPGDPPCWT 1050
Db 1994 -----QPKGQGOQOQGOQGOQGOQGOQGOQGOQGOQGOQGOQGOQGOQGOQGO 2039
Qy 1051 SGLPFPVPREVIKASHPADPSAFY-----APP---GHPLPLGLHDTARPVLPR 1098
Db 2040 -----ERKELTIVREVRQDPGLKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 2090
Qy 1099 PPTISNPPLISSAKHPSVLEROIGALSQGMVQLHVPVSEHAKAPGVPTMGLPLMDP 1158
Db 2091 PAAVAPPPP-----AHP-LTFTSTGCAGSNGTSDSLATLSVNSVMGMVGIGHPGMAH 2144
Qy 1159 KKLAPFGVKQEOQLSP--RGAGPPSELGVPTAQEASVL-RGTALGVPVSGISITKGIPT 1215
Db 2145 ASSAGGIGVDKATITPVVKSSSGSSKSGGSSASSHSTATPETIINYVVAHPQRGILPP- 2203
Qy 1216 RVPDSDAI--TYRGSITHGTADVLKGTITRI--IGEDSPSLDRGRSDSLPKGHVYIEG 1272
Db 2204 --PSQSHVHPAHSHTQHPAHPQSHSHGHQTLQVPEPEPOTLDLSIKKPPRDCGSPHTG 2261
Qy 1273 KKGHLVSEYGMVSTQCKEDGRSS---GPPHETAAPK---RTYDMMEGRVCAISSA 1325
Db 2262 -----AGGSSSGSGSGGSPSSDRHHGPPPTTMSMKHIVRSGGMYRGDTVTVPVSLA 2312

Qy 1326 -----SIEGLMGRAIPPERHSPHLKEQHIIHGSIITQGIIPRSYVEAQ-----ED 1369
Db 2313 APSSYLPTRSVKTIIGGGVVP-----GVLPGVGFSALYLQVPVPVPI 2356
Qy 1370 YLRREAKLLKREGTPPP--PPPSRDLTEAYKTQALGPLKLP--AHEGLVATVKEAGRSI 1425
Db 2357 SISQOGLPFRAGOPPPPAQPPSGRGVAK-----PPKLSPOQAHLHLHGHSPPSOQO 2409
Qy 1426 HEIPREELRHTPELPLAPRPL--KEGSITQCTPLKYDTGASTTGSKKHDVSLIGSPGT 1483
Db 2410 QO 2465
Qy 1484 FPPVHPLDVMADARALERACEEELKS-RPGTASSG--GSIARGAPVIVP----- 1531
Db 2466 LSPKFD-----GLVRQTTPEGVGSGVPGGASGSGKHGSIITQCTPLHMPHLESKRP 2517
Qy 1532 -----ELGKPRQ--SPLYEDHG-----APFAG----- 1552
Db 2518 YESYKSSQRHSPAQQPGNQQLPPPPQSSPQAPPQPGYGVGVSSPYARSPPFAGVBPQ 2577
Qy 1553 -----HLPRGSPVT-----MREPTPL 1569
Db 2578 QVLSTROIWMHDYITSQOMQO 2637
Qy 1570 QEGSLS-----SSKASQDRKLTSTPREIAKSPHSTVP-----BHH 1604
Db 2638 SVGSASGFAYGDKESAPRGREYSSRASPADHVNSTP---SPHRTPPQRQGVQRH 2692
Qy 1605 -----PHP-----ISPYEHLRGVSDVLYRSHIPLAFDPTSTP-----RG1PL 1643
Db 2693 NTGSKPPSPAAPPSPRMHMPYQY---APSGHDALASFVDVAVQOQPLVPSPKDDKSPG 2749
Qy 1644 DAAAYVYLPRH--LAPNPTVPHLY-----PP--YLIRGYPDTAALENQTIINDVIT 1691
Db 2750 PSTAPGVQSGPPLGSPPLPPHVAVGVAQPPPTAHHDQRYDLTLHHHHHTLVQOOQI- 2808
Qy 1692 SQOMHNTATAMARADMLRGLSPRESSLALNYAAGPRGIIDLQVPHPLVLVPTPGTP 1751
Db 2809 AQOQHYSRLNVAQ--VDMQRMQAKRVMRHHQHQVQOQOQOQOQOQOQOQOQOQOQOQO 2857
Qy 1752 ATAMDRLAYLPTAPQPFSSRRHSSPLSPGGPTHUTKPTTTSSRRDRDRDRDRDRDRDR 1811
Db 2858 ALERDR-----EMQERMERERERE-----RERERERERERERERERERERE 2900
Qy 1812 KSILTSITTV-----EHAPIWRPCTEOSGS--SGSSGGGSSSSRRPASHAHQHSPLS 1864
Db 2901 RREQDARRVVAERERHDS--RMRMRFAGNVVTGSGAGGGGPS-PQOFLRA----- 2950
Qy 1865 PRTQDALQORPSVLHNTGMKGIIITAVEPSPKPTVLRLSTSSPVRPAATFPFATHCPLGCT 1924
Db 2951 -----SVPETGPPRSIPDRERESVYRQAHGAPAPEDTP--GQ 2985
Qy 1925 LDGVVYPLMBPVL---LPKEAPRVARPER--PRADTGHAFIAKPPARS-----GLEPASS 1975
Db 2986 LSA--QSLIDAIKHEINRSDNATAGPREPPSPFVHAPL--PPRGSGSGGTGTTRSP 3041
Qy 1976 SKGSEP---RPLVPPVSGHATIAKPAKLAHPHASPDPPA-----PPASADP 2021
Db 3042 ANVLHPWYLDLROPLDGGAGSMLTAENNGKP--SSSGSPSVINIDLDQERISAAAAVA 3099
Qy 2022 HREKTQSKPFSIQLELRSLGYHGSYSPEGEVPSVSSPSLTHDKGLPKHLELDKSH 2081
Db 3100 QO 3154
Qy 2082 LEGELRPKQPGVKLGGEAAHPLHLP--LPSQPSSSP-----LLOTAPGVKG 2128
Db 3155 ITSDYGTN-----PHLRPPYMAVLOETQSILPPDRWKQNRMQKAEAKH 3200
Qy 2129 H---ORVVTILAOHI SEVITODYTRHHPOQLSAPLAP-LYSFPG----- 2168
Db 3201 HSQO 3260


```

Db 3385 LGAVGGVSLVGGSG-----GIAGG--PGGVS----- 3411
Qy 2455 SSAGSTFPYPLIMRLQAGVMAAPPPGCPAGSGPLA-----GPHH----- 2496
Db 3412 -----VG------GVPGGGPGSGGGGGGGHNSSSQASAAVAAA 3446
Qy 2497 --ANDEPKPLCSQYETLSDSE 2517
Db 3447 VAAAASESKPLLSKDALSDSD 3469

RESULT 16
Q7ZUE9 PRELIMINARY; PRT; 302 AA.
AC 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to nuclear receptor co-repressor 1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Body;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049302; AAH49302.1; --
KW GO; GO:0004872; F:receptor activity; IEA.
SQ SEQUENCE 302 AA; 36126 MW; 8B82177688970BC6 CRC64;

Query Match 6.3%; Score 827.5; DB 13; Length 302;
Best Local Similarity 40.6%; Pred. No. 2.1e-32;
Matches 181; Conservative 40; Mismatches 66; Indels 159; Gaps 6;

Qy 11 TWRATEPRYPHSLSYVQIARTHTDVGLLEYQHHSDYASHLSPGSIQORRRPSSL 70
Db 12 SFSESGSRYPHSLSYVQIARTHTDVGLLEYQHHSDYASHLSPGSIQORRRPSSL 59
Qy 71 EFOGNERQELHLRSHSYLPEL-----GKSEMEFIESKRPRLELLPDLRLPSPILLAT 126
Db 60 EPHPGTERPPE-----RRHGYEQQFHAISAQOEQALESKRPRIE----- 99
Qy 127 GQPAGESEDLTKRSLTGKLEPVSPSPPPHTDPELELVPPRLSKEELIQNMORVDREITWV 186
Db 100 -----NVS 102
Qy 187 EQQISKLKKQOOLEEAAKPEPEKVPSPPTIESKHSLSVLQIYDENKKAHAHRILE 246
Db 103 ETHEFSR----- 108
Qy 247 GLGQVZELPLYNOPSDTRYHENIKINOAMRKKLILYFKRRNHARKQKFCQRYDQLM 306
Db 109 -----SQSDTKVTHDNIKNQVNRKLLILFFKERNHARKQKFCQRYDQLM 157
Qy 307 EALEKKVERIENPPRRAKESKVREYKEQFPEIRKQRELQERNQSRVQSGLSMSAA 366
Db 158 TEWEKKVERMENPPRRAKESKVREYKEQFPEIRKQRELQERNQSRVQSGLSMSAA 216
Qy 367 RSEHVSHIIDGLSEQENLEKQMLAVIPPMYLDADQORIKFTNNMGLMADPMKYKDR 426
Db 217 RSEHSHIIDGLSEQENLEKQMLAVIPPMYLDADQORIKFTNNMGLMADPMKYKDR 276
Qy 427 QVMNWSQEXETPREKFMQHPKNGF 452
Db 277 QVMNWSQEXETPREKFMQHPKNGF 302

RESULT 17
Q86YY2

```

```

ID Q86YY2 PRELIMINARY; PRT; 453 AA.
AC Q86YY2;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE NCOR isoform d.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yu L.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF303584; AAC32940.1; -- ADC3D56F36EADB20 CRC64;
SQ SEQUENCE 453 AA; 49417 MW; ADC3D56F36EADB20 CRC64;

Query Match 4.9%; Score 650; DB 4; Length 453;
Best Local Similarity 39.9%; Pred. No. 1.3e-23;
Matches 191; Conservative 60; Mismatches 145; Indels 68; Gaps 22;

Qy 2103 LPHLRPLPESQSSPLLTQAPG-----VKGHQVVTTLAQHISEVITQDYTRHH--- 2151
Db 28 LHHYRP-----QGESPPQQLPSSQAEQGVPRTHRLITLADHICQIITQDFARNQVSS 84
Qy 2152 --PQQLSAPLPAPLYSFFGA--SCPVLDLRRPPSLYLPDPD-----HGAP-ARGSPH-- 2199
Db 85 QTFPQ-----PPTSTFQNSPSALVSTPV--RTKTSNRYSPESQASVHHQPGSRVSPENL 138
Qy 2200 ---SEGKRSPEPNKTSVLGGEDGIEPVSPPEGWTEPGHRSRAVPLLYRDGEQTEPSR 2256
Db 139 VDKSRGSRFGKSPERSHV---SSEPYEPLSPQ--VPVVHEKQDSLILLLSQGAEPAEQR 193
Qy 2257 MGSKSPGNTSQPPAFPSKLTESNAAVSKKQEIKNKLTNRNPEFYNISOPGTEIFNM 2316
Db 194 NDARSFGSISLPSFTKL-ENTSPVWKSKQEIFKLNSSGGSDMAAAGQTEIFNL 252
Qy 2317 PAITGTGLMYRSQAOEHAHSTNMLEAIRKALMGKYDQWEE-----SPPLSANAFNP 2370
Db 253 PAVTTSVSSSRGSHSPADPAS-NLGLDIIRKALMGSDKVEDHGVMSQPMGV---VP 308
Qy 2371 LNASASLPAAMPITAAAGSDHTLTSPGGG-GKAKVSGRPSRKAQSPAP--GLASGDRP 2427
Db 309 GTANTVS-----VSGETREEDGDPHSGVCKPKLISKNSRKSXSPIPGQYLGTERP 363
Qy 2428 PSVSSHSEGDGCMRRPTLTNRVWEDRPSSAGSTFPFPNPLIMELQAGVMAAPPPGCPAG 2487
Db 364 SSVSSHSEGDYHROTP--GWAWEDRPSSTGTQFPYNPLTMRM---LSSTPTPTIACAP 418
Qy 2488 SG-PLAGPH--HAWDEEPKLLCSQYETLSDSE 2517
Db 419 SAVNQAAPHQONRIWEREPAPLLSAQYETLSDSD 452

RESULT 18
Q8BK32 PRELIMINARY; PRT; 291 AA.
AC Q8BK32;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Weakly similar to BB329D4.2.1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;

```


QY 795 EPTFASBATGAPTPPPAPPSPAPPVPPVVEKEEETAAAPPVEGEEOKPPAAABELAVD 854
 Db 2418 QVUSMTETIREDADQMPKPSOA-----ESRRESIAESIKASSPRDEKSPPLASKEASR 2469
 QY 855 TGKAEPPVK-----SECTEABEGPAKGDAEAAEAETAGALK-AEKKEGSGR 902
 Db 2470 PGSAVESIKYDLKPFQIKKODKSTEHSHRRESLEDKSAVTSEKSVSRPLSVASDHAAVAL 2529
 QY 903 ATTAKSGAPOSD-----SSATCSADEVDEAE-----GDKNRLSP 940
 Db 2530 EDDAKGSLSPKDKRPGFVAETVSPIEEATMEFSEKIEVVEKSSALSLQGGSGGKLQTD 2589
 QY 941 RPSLLTPTGD-PRANASQ-----KPLDLKQLKQRAAIPPIQVTKVHEPPREDAAPT 992
 Db 2590 SSPVDVABEGDFSHAVASVSTVTLTKPAELAQ-----IGAATVSSPLDEALRT 2639
 QY 993 KPAP-----PAPPPONLOPESDAPQOPGSSPRGKSRPAPPADKEAPAAEAQKLP 1044
 Db 2640 PSAPHEHISRADSPAECASEEIASQKSPQVLKESR-----PAWAEKDDAQLKSSVE 2694
 QY 1045 DPCCWTSGLPPVPVPPREVIK-----ASPHAPDPSAFSVYAPPGHPLPLGLHDTARPVL 1096
 Db 2695 D-----LRSPVASTEISRPAAGETASSPIEAPKQDFA-----EFEQAEKAVL 2737
 QY 1097 P-----RPPTISNP-----PPLISSAKHPSVLEROIGATISQGSVQLHVPYSEHA 1141
 Db 2738 PLTIBELKGNLPLTSSFPDVHAGDFPQTSTPTSSPTVASVQPAELSK-----VDIEKTA 2790
 QY 1142 KAPVGPV-----TWGLPLPMDPKKLPFGVKQEOQLSPRQAGPPRESLGVPTAQEA----- 1192
 Db 2791 SSPIDEAPKSLIGCAEERPEPSPASAKDAEAEVSKDASRPSPVVESTKADTKGDIS 2850
 QY 1193 ----SVLRGTALGSPVGGSIKIGIPSTRVPDSATYRGSTHTGTPADVLYKGTITRIIG 1248
 Db 2851 PPSPEVLEG-----PKDDVEKSKSRPPSVSA-----SITGDSKDVSRPASVSVK 2899
 QY 1249 EDSPSRLDR-----GREDSLPGHVIYEGKGHVLSEYEGM--SVTQCSKE 1292
 Db 2900 DEHDKAESRRSIAKVESVIDAGKSDSKSSQDSQKOEKSTLAKESARRESVVESSKD 2959
 QY 1293 DG-RSSSGP-----PHETAAP-----KRTYDME----- 1315
 Db 2960 DAEKSESRPESVIASCEPVRESKSPDKDTSRPGSMVESVTADEKSEQOOSRESVAE 3019
 QY 1316 -----GRVGRALISSASIEGLMGRAPPPERHSPHILKEOHIRGISTQIGIPRSYVEA 1366
 Db 3020 SVKADTKDQKSOEASRPSVDELLKDDDEKOEKRRQSIKSGHKAMSTGDESMDKADK 3079
 QY 1367 QEDYLRE--AKLKEGTTPPPPP-----SRDLTEAVYKQALGPLKLKPAHE 1412
 Db 3080 SKPSPRPESVAESIKHENTKOEKSPGSRSDSVAESIKSDITKGEKSP--PSKEVSRPE 3137
 QY 1413 GLVATVK-EAGRSIHEIPRELR--HTPELPLAPRPLKEGSIQGT--PLKYDTGASTTG 1467
 Db 3138 SVVGSIKDEKAEARRESVAESVPESSKDATSAP-PSKEHRSPEVLSGLK-DEGDKTIS 3195
 QY 1468 SKK-----HDVSLIGSPGRTPPPVHPDVMADARALACVYESIKSRPGPTASSGGS 1521
 Db 3196 RRVSVADSIKDEKSLVQSASRPE-----SEAESLKDAAPSOETSRPESVTE---S 3245
 QY 1522 IARG-APVIVPELGPQSLTYEDHGAGFAGHLPGRGSPVTMEPTPRLOEGLSSSK-- 1578
 Db 3246 VKDGKSPVASKEASRPASVAENAKDSADESKEQRP-----ESLPQKAGSIKDEKSP 3297
 QY 1579 -ASQD-----RKUTSTPREIAKSPHSTVPEHHPHTIPSPVEHLIRGSGVDLYRSHIPL----- 1630
 Db 3298 LASKDEAEKSKESRRRESVAEQFPPLVSKSVRSRPASVAESV---KDEAEKSKESPLMSXE 3354
 QY 1631 AFDPTSIPIRGPLDA-----AAAAYTLPHLIAPNPTYPHLYPPYLIRGYPDATA 1678
 Db 3355 ASRPASVAGSVKDEAEKSKESRRRESVAESKSPPLPSKEASRPA----- 3396

QY 1679 LENRQTIINDYITSQOMHINTATATAQRAADMLRGLSPRESSLAINYAAGPRGIIDLQVP 1738
 Db 3397 -----SVAESVKDEADKSKESRRRESGAESKPLASKEA----- 3429
 QY 1739 HLPVLVPTPTPTATAMDRLAYLPTAQPFSSRRSSPLSPGPGTHLTTPKTTTSSS--E 1795
 Db 3430 -----SRPASVAESI KDEAEKSKESRRRESVAESKSPPLPSKEASRPTSVAESVKDE 3479
 QY 1796 REEDRDRDRDREREKILSTTTTVEHAPITWRPCTEQSSGSSGSSGGGGSSRRPASHS 1855
 Db 3480 AEKSKESR-RDSVAESKPLASKEASRPASVAESVQDEAEKSKESRRRESVAESKPLASK 3538
 QY 1856 HAHQHSPIPTQDALQ--QRPSVLHNTGMGIITAVEPSKPTVL-----RSTSTS 1904
 Db 3539 EASRPASVAESIKDEAEKSKESRRRESVAESKPLASKEASRPTSVAESVKDEAEKSKES 3598
 QY 1905 SPVRAATFPFATHCPLGGLTLDGVVPTLMPEVLLPKAPRVARPERPRADTGHAFKAP 1964
 Db 3599 SRDSVAESKSP-----LASKEASRPASVAESVQD--EAEKSKEE 3634
 QY 1965 ARSGLEPASSP---SKGSEPREPLVPVSGHATARTPAKNLAPHASP---DPPAPPASA 2018
 Db 3635 SRRESVAESKPLASKEASRPASVAESVKDDAEKSKESRRRESVAESKPLASKEASRPASV 3694
 QY 2019 SDPHREKTSQKFSQLELRSLGYHGSYSPEGVEPVSPVSSPSLTHDKGLPKHLEB-- 2076
 Db 3695 AESVKDEAEK---SKEESRRRESVAESKSPPLPSKEASRPTSVAESVKDEAEKSKESRESV 3751
 QY 2077 LDKSHLEGRLPKQPGVKLGGEAAHLPHLRPLPESOPSSPPLQTPAGVKGHQVWTLA 2136
 Db 3752 AEKSKSLASKEASRPASVAESVKDEAEKSKESRRRESVAESKPLAS-----KEASRPASVA 3806
 QY 2137 QHISEVI--TQDYTRHHPOOLSAPLPAPLYSPFGASCVPDLRRPPSDLYLPPDHGAPA 2194
 Db 3807 ESKVDEAEKSKESRRRESVAESKSPLSK-----EASRPTSVAESVKDEADKSK 3854
 QY 2195 RGFSEHGEKGRSPENKTSVLGGGEGDGI1BPVPPGEMTEPGHRSASVYLLYRDGEQTEP 2254
 Db 3855 EESRRRESGAESKPLASM-----EASRPTSVAESVK-----DETEK 3889
 QY 2255 SRMGSKSPGNTSOPPAFFSKLTESNAMYKSKQEIKNKLNTHNNEPEYNIQOTETFP 2314
 Db 3890 SKESRRRESVTEKSP-LPSKEASRPTSVAESVKDEAEKSKESRRRESVAESKPLASKS 3947
 QY 2315 NMPAITGTGLMTRSAQVQEHASTNMCLEAIIRKALMGKYDOWEESPPLSANAFN---PL 2371
 Db 3948 SRPASVA-----ESIKDEAGTKOESRRRESWSPESGKAESIKGDQ 3986
 QY 2372 NASASLPAAMPITAAAGRSDHITLTSFGG--GKAKVSGRP-----SSRKAESKPLASGD 2425
 Db 3987 SSLASKETSPPDSVRESVKDETEKPEGSAIDKQSVASRPESVAVSAKDEKSP-----HS 4041
 QY 2426 RPPSVS--SVHSEGDCCNRTPLTNRV---WEDRPSSAGSTPPFPVNPMLRLQAGVWASPP 2480
 Db 4042 RPESVADKSPDASKEASRSLSVAETASSPIEEGPRSTADLSLPLN-----LTGEA 4091
 QY 2481 PPGPAGSGPLAGPHHAW-----DEEPKPLICSQ 2509
 Db 4092 KGKLTSSPIDVAGDFLEKVAESSRPAVLK 4125

RESULT 21
 076891 PRELIMINARY; PRT; 5327 AA.
 ID 076891
 AC 076891;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE EG:49E4.1 protein.
 GN FUTSCH OR EG:49E4.1 OR CG3064.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;


```
QY 2361 PPLSAN-----AFNPL-----NASASL----- 2377
Db 1812 PPTADTALQSEYARPHVAFSGVEQVYHHPMGMYRERLEETKNAQAAAQSRSL 1871
QY 2378 -----PAAMPITTAADGRSDHTLTSPGGGKAKVSGRPSRKAQSPAGLASG 2424
Db 1872 DPHWMEYRRGIHPSQFPLVYANPAISQMERELGIPPHVGLDPGSHVMRMPQPEAGF 1931
QY 2425 DRPPSVSVHSEGCNRRRTPLTNRWEDRPSSAGSTPPFPYNPLIMRLOA 2473
Db 1932 QLPENVG-----QYPRNMLIPREPHSDVLLRMSYADQLQYLOA 1970

RESULT 24
Q9BZS0 ID Q9BZS0 PRELIMINARY; PRT; 2406 AA.
AC Q9BZS0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Kappa B and v(D)J recombination signal sequences binding protein.
GN KRC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Thymocytes;
RC MEDLINE=2110080; PubMed=11161801;
RA Hicar M.D., Liu Y., Allen C.E., Wu L.C.;
RT "Structure of the Human Zinc Finger Protein HIVEP3: Molecular Cloning,
RT Expression, Exon-Intron Structure, and Comparison with Paralogous
RT Genes HIVEP1 and HIVEP2."
RL Genomics 71:89-100(2001).
DR EMBL; AF278765; AAK01082.1; -.
DR HSSP; P15822; 1BBO.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003700; P:transcription factor activity; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR InterPro; IPR002034; AIPM/HcIt_synth.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 5.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 2406 AA; 259381 MW; 7F498F2BBF3AE93F CRC64;

Query Match 4.2%; Score 552.5; DB 4; Length 2406;
Best Local Similarity 19.8%; Pred. No. 5.2e-18;
Matches 553; Conservative 323; Mismatches 966; Indels 953; Gaps 136;

QY 3 GSTQLVAQTRATEPRYPHSLSYVQIARTHTDVGLEYQHRSDYASHLSPGS---II 59
Db 69 GSQKGTGQ--QKPKRPPIEASVHI-----SHVQPHLTPAFMSPGKPEHLEGSTWQLV 122
QY 60 QPQRRRP--SLLSEFPQG-NERSQELHLPESHVLP-----ELGKS 98
Db 123 SPMRLGPGSGSLIA---PGLHPSQ---LLPSHASIIPPEDLPGVKFVPRPSQVSLKPT 176
QY 99 EMEFTESKRPR-----LELLPDLRLRPSLLA-----TGO-----PAGESDITKD--- 138
Db 177 EEAHKERKPKQPGKIQYQCSRCPKAFSVLQKHRSHTGERPPYCGCGSFYTKSNLY 236
QY 139 -----RSLTGKLEPVSPSPPTDPELELVPPRLSKBELIQNMQRVDREITWVEQOI 190
Db 237 KHRKSHAHRIKAGLASGGMGYPH-GLEMERIPGEEFEEP-----TEGESDSEET 288
QY 191 SKLKKKQOOLEEAAKPEPEKVPSPPIES---KHSLSVQIYIDENRKAABAAHRLLEGL 248
Db 289 SATSGHPAELS-----PRPKQPLSSGLYSGSHSSHERCSLSQSSSTAQSLEDPPFPV 342
```

```
QY 249 GPOVELPLYNQSDTRQYHENIKINQAMKKLI-----LYFKRRNHARK 292
Db 343 ERSSEHPLSHKEDTHTIQKLAHLRLSERKKVIDEQAFLSPGSKGSTGYFSRSAAEQ 402
QY 293 QWKQKFCQYDQIMEALEKKVERIENNRRAKESKVRREYKQFPEIKQRELQERMOS 352
Db 403 Q-----VSPNTNAK-----SYAEIIFGKC-----G 423
QY 353 RVCGSGSLMSAAR-----SEHEVS-----EIIDGLSEQENLEKQMLQAVIPP 397
Db 424 RIQORTAMLTATSTQPLPLSTEDKESLVPLSVPRTOVIEHTKLITINE-----AVVDT 478
QY 398 MLYDADQOQRIKFINNMGLMADPMKYKDRQVMNMWSEQEKETFREKPMQHPKNGFIASF 457
Db 479 SEIDSVKPRRSLSRSSMESP-KSSLYREPLSSHSEKTKPEQSLLSLQHPPTAPPVPL 537
QY 458 LERKTV--AECVL-----YYLTKKNYKSLVRSY-----RRGKSQOQOQOQO 502
Db 538 LRSHMPSAACTISTPHHPFRGSYFDDHITDSEALSRSHVFTSHPRMLKQPQAIELPL 597
QY 503 QOQOQOQOQMPRSSQOEKDEKEKEAEKEEKEPEVENDKEDLLKEKTDGTSGEDNDEKE 562
Db 598 GGEYSSEEPGPS--KDTASKPSD-----EVE-PKESLTKT----- 632
QY 563 AVASKGRKTANSQRRKG-----RITRSMANEANSEAITPQOSAEALASLMELNESSRWTE 617
Db 633 ---KGLKT-----KGVIECNICGARYKRDNYEAHKYCYSELQIAKPIASGTHTS 682
QY 618 EEMETAKGLLEHGRNWSAIAIRVMGKTVSQCKNFYNYKKRQNLDEILQOHLKWEKER 677
Db 683 PE---AEKSQIER-EPWSQ-----MMHYKLGTTLEL 709
QY 678 NARRKKKAPAAASERAAPVVEDEMEASGVSGNEEMVEAEALHASGNEVPRGEC 737
Db 710 TPLRKRKEKSLGDEE--PPAFESTKSPGS-----P 740
QY 738 GPATVNNSSDITESIPSPHTEAADTQONGPKPATLGADGPPPGPTTP-----PRRT 789
Db 741 GPSDAARNLPLESTKSP-AEPSKSV-----PSLEGPTGFQRTPKPGSGESGKERRT 792
QY 790 SR---APIETPASEATGA-----PTPPAP-----PS 814
Db 793 TSKEISVIQHTSSFEKSDSLEQSGLEGEDKPLAQFPSPPPAPHGSAHSLQKLVQPN 852
QY 815 PSAPPPVVPKEEKEETAAAPPVEEGEEOQPPAAAEELAVDTGKAEPEPKSECTEE--AAE 872
Db 853 IQVPEILVTEEPDRPDTEPEPPPEKPEK-----TEEFQWQPSQTLAQLPAEK 900
QY 873 GPAGKD---AEAATAEGALKAEKKEGGSGRATTAKSGAPQDSDSSATCSADEVDEA 929
Db 901 APPKKRLRLAEMAQSSGESSFESSVP---LSRSPQESNVSLSGSSRSASFERRDHGKA 957
QY 930 EGGDKNLLSPRP-----SLLTPTGDP-----RANASPKP-----LKLKOLKORAAAIPP 975
Db 958 EADPDSMDMRPKPLGTHMLTVPSHPHAREMRRSASEQSPNVSHSAHMTTRKSPDYGS 1017
QY 976 IQVTKVHEPPREDAAATKPAAPPAPP-----PPQNLQ-----PESD---AP--QQ 1014
Db 1018 LSLT-----GPSAPAPVAPPAGEAPERRKCFLVRSPSLRPPESSELEVA PKGRQ 1067
QY 1015 PGSSPRGKSRPAPPADEKFAAEQAQLPGDPPCWTSGLPFP-----VPREVITKA----- 1065
Db 1068 ESEBPQSSSKPSAKSSLSQISSAATSHGGPGGKPGQDRPALGPTVPVTEALQVFFHP 1127
QY 1066 -----SPHAPP--SAFSVA-----PPGH-----PLPLGLHD 1090
Db 1128 VAQPTLHEKPYLPVPVSLFSFQHLVQHEPGQSPFFFTQAMSSLLSPFNMPLPPSLFQ 1187
QY 1091 TARPVLPRPTISNP-----PPLISSAKHPSVLERQI---GAISQGM----- 1129
Db 1188 A--PPLPLQPTVLHPGLHLPLQMLPHPANIPFRQPPSFLPMPYPTSSALSSGFFLPQSQ 1245
```



```
Qy 1130 -----SVQLHVPYSEHAKAPVGVNTWGLPLPMDPKKLAPFSGVKQB-QLSPRGQAGPP 1181
Db 1246 FALQPLGDSVHLPQIKTSLAPATGSAG-----LSPQSYSSDIRLPP---VAPP 1293
Qy 1182 ESLGVPTAQEASVLRGTALGVPGSGITKGIP---STRVPSDSALTIRG-----SITHG 1232
Db 1294 ASSAPTSAPP-----LALPACPDPMVSLVFPVRVQTNMPSYSGSAMYTTLSQIILVTSQG 1348
Qy 1233 TPADVLY-----KGTITRIIGED-----SPSRLDRGREDLSPKGH-----VIYEGKK 1274
Db 1349 SSAITVALPKPREPSKG---TIVCGADVHEVGFSGSLSEESQSRAPPTPLVAVPTLPERK 1406
Qy 1275 GHVLSYEGGMVTOCSKEDGRSSGPPHETAAPKRTYDM-----MEGRVGRAISASIEG 1329
Db 1407 GTSLSSESILSL-----EGSSSTAGGSKRVLSPAGSLELTMETQOQKVEEESKADEK 1461
Qy 1330 LMGRAIPP-----ERHSPHLKEQHHRGSIITOGIPRSVVEAQEDYLBREAKLL 1378
Db 1462 L--ELVKFCSVVLSTEDGKREPKSHL-----GNQGG-----RELEML 1499
Qy 1379 K-----REGTTPPPPSRDLTEAYKTOALGPLKL---KPAHEGLVATVKEAGRSIHE 1427
Db 1500 SSLSDSPSDTKEIPLPLPALSHGQAPGSEALKYQPSGKPHRGLTPL-----S 1550
Qy 1428 IPRELHRTPELP-LAP---RPLKEGSIQTGTPKYDTGASTGSKKHDRSLIGSPORT 1483
Db 1551 VKKEDSKQEPDPLSLAPPSLPLSE---TSRPAKSEQG---TDSKK-----VLQ 1594
Qy 1484 PPPVH-----PLDVM-ADARALERA---CYEESLKSRPPTASSGSGSIARGA 1526
Db 1595 FPSLHTTTNVGWCVLNYIKPHNIQHADRSSVYAGWCISLYNPLPGVSTKAALSLLR-- 1652
Qy 1527 PVIVPELCKPROSLTYEDHGAPFAGHPLRGSPTVMTREPTPELOB---GSLSSSKASODR 1583
Db 1553 -----SKQVSKETYMTATAP---HPEAGRLVPSSSRKPRMTVEHPLSVSPGQXDL 1702
Qy 1584 KLTSTPREIAKSPHSTVP---EHPHPITSPYEHLLRGVGVLDYRSHIPLAP----- 1632
Db 1703 ARVEKEEERRGEPEEDAPASQGEFARIKIFE-----GG---YKSNEYVVVRGRGK 1753
Qy 1633 -----DPTSIPRGIPLDAAAAYLPRHL-----APNPTYPHLYPYLRGYDPT 1676
Db 1754 YVCBEGCIRCKPMSLKKHIRTHTDVRPYVCKGHCFAPKTKGNLTGHMKSKAHKKCOET 1813
Qy 1677 AALENRQTIINDYITSQMHNTATAMAQRADMLRGLSPRESSLALNVAAGPRGIIDLSQ 1736
Db 1814 GVLE-----ELEABEGTS---DDLQDSEGRGSEAV----- 1842
Qy 1737 VPHLPVLVPTPGTATAMDRLAYLPTAPQPFSSRRHSSPLSPGCPHTLTKPTTTSSSR 1796
Db 1843 -----BEHQFSDL-----EDSDS 1855
Qy 1797 ERDRDRDRDREREKSLTSTTT-----VEHAPIWRP-----GTEQSSG 1836
Db 1856 DSDLDEDEDEEESQDELSRPSSEAPPPGPHALRADSSPLGQPQPDAPASGTEATRG 1915
Qy 1837 SSGS-----SGGGGGSSSRPA-----SHSAHOHSPISTR----- 1866
Db 1916 SSVSEAEELTASSCSMSSQSPHGLPWLGAPLGSVEKDTGSALYKVPSPRRPWSPSKEA 1975
Qy 1867 -TQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSPV----- 1907
Db 1976 GSRPPLARKHSLTKNDS-----SPQRCSPAREPOASAFSPGLHVDPCGMGPLPCG 2027
Qy 1908 RPAATFPFPAHCPLG-----GTLDGVYPTLMEPVLLPKCAP-----RVA 1946
Db 2028 SPRLQLSPLTLCPLGRELAPRAHVLSKLEGTDD---PG-----LPYSPTRRMSPGOAE 2078
Qy 1947 RPERPRADTGHAFIAPKPPARSGLSPASSPSKGSE-----PRPLVPPVS-CHATIARTP 1998
Db 2079 SPPR-----SAPGKVALAGPGSPAGEHGFGLGAPRVLFPPAPLPHKLLSKSP 2128
Qy 1999 AKNLAPHASPDPPAPPASADPHREKTQSKPFSIQLELRLSLGYHGSSYSPGVEPVSP 2058
```

```
Db 2129 ETCSAPWQKA-ESRSPSCSPGAH--PLSSRPFS-----ALHDFPHGHLARTENI----- 2176
Qy 2059 VSSPSLTHDKLGPKHLELDKSHLEGLRPKQGP-VKLGCEAAHLPHLRPLPESQPS 2117
Db 2177 -----FSH---LPLHSHLTRA-----PCPLPIGG-----IQMVQARPOAH 2210
Qy 2118 PLLQATP-----GVKGHQRVVVTLAQHISEVITQDYTRHHPOQLSAPLPAP---LYSPFG 2168
Db 2211 PTLPLPGPTAAWVSGFSGGSDLTGARE-----AQERGRWSPTESSASVSPVAKVSKFTL 2265
Qy 2169 ASCPVILDRPPSLYLPDPDHGAPARGSPHSEGGKSPENKTSVLGGEDGIEPVSP 2228
Db 2266 SS--ELEGRDYPKERERTGGGPGRPDPWTGHGTGAPAEPTTHS-----PCTPP 2312
Qy 2229 EGMPTEPGHRSADVPLLYVRDGEQTEPSRMGSKSPGN---TSQPPAFFSK---LTESNSA 2281
Db 2313 DTLPRPPQGR-----RAAQSWSPLESPPRANPEPSATPPLDRSSVGCCLAEA-SA 2363
Qy 2282 MVKSKQOEINKKLNTH-----NRNEPEYNISQP 2309
Db 2364 RFPARTNLGSESTRQDSPKPSGSGEPRAHPHP 2398

RESULT 25
Q9NHX6
ID Q9NHX6 PRELIMINARY; PRT; 1966 AA.
AC Q9NHX6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GRUNGE.
GN GUG OR CG6964.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Erkner A., Roue A., Core N., Angelats C., Vola C., Fasano L.,
RA Kerridge S.;
RT "Grunge is required for proximal and ventral leg development in
RL Submitted (DSC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR EMBL; AF217844; AAF34752.1; -.
DR Flybase; FBgn0010825; Gug.
DR GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR000949; ELM2
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF03154; Atrophin-1; 2.
DR Pfam; PF01448; ELM2; 1.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00717; SANT; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 1966 AA; 208033 MW; 96AF90E2082E770C CRC64;

Query Match 4.2%; Score 552; DB 5; Length 1966;
Best Local Similarity 20.0%; Pred. No. 4.3e-18;
Matches 490; Conservative 235; Mismatches 834; Indels 888; Gaps 111;

Qy 328 KVRYYEK-----QFPEIRKORELOERMQR-----VCQSGLSMSAARSEHVESE 374
Db 16 QVNDVYAKLPDYNPISSFP-IDKETDERLEESRWSPGWADGDLMLFLRAARSMAAFQG 74
Qy 375 IIDGLSQENL-----EKQMRQLAVIPMLYDADQORIKFINMGLMADPMKVYKDRVM 429
Db 75 MCDGLEDDGCLAAASRDTPINALVDHDSGVDPKA-----LQALVKCPVSKGIDKK-- 126
Qy 430 NMWSEKETPREKFMQHPKNFGLI-ASFLEKRTVAECVLYYYLTCKNENYKSLVRSYR 488
```



```
Qy 2361 PPLSAN-----AFNPL-----NASASL-----PAAMPI 2383
Db 1812 PPYADTALQSEYARPHVAFRELEIKNAQAAASQSRDLDPHMEYRRGIHPSQFPL 1871
Qy 2384 TAADGRSDHTLTSFGGKAKVGRPSRKAKSPAPGLASDRPSV 2430
Db 1872 YANPAISQMERERLIGPPPHVGLDPGSHVMRMPQPPEAGFQLPPNV 1918

RESULT 26
Q8T9N4 PRELIMINARY; PRT; 1985 AA.
ID Q8T9N4
AC Q8T9N4;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Transcriptional corepressor Atro.
GN GUG OR ATRO OR C6964.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=21652534; PubMed=11792320;
RX Zhang S., Xu L., Lee J., Xu T.;
RA "Drosophila atrophin homolog functions as a transcriptional
RT corepressor in multiple developmental processes.";
RL Cell 108:45-56(2002).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR EMBL: AF475087; AAL78679.1; -.
DR FlyBase; FBgn010825; GUG.
DR GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR000949; ELM2.
DR InterPro; IPR001005; Myb DNA binding.
DR Pfam; PF03154; Atrophin-1; 2.
DR Pfam; PF01448; ELM2; 1.
DR Pfam; PF00249; myb DNA-binding; 1.
DR SMART; SM00717; SANT; 1.
DR DNA-binding; Nuclear protein.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 1985 AA; 210423 MW; 120A78F6C8FD67F CRC64;

Query March 4.2%; Score 552; DB 5; Length 1985;
Best Local Similarity 20.0%; Pred. No. 4.4e-18;
Matches 505; Conservative 242; Mismatches 855; Indels 920; Gaps 117;

Qy 328 KREYVEK-----OPPEIRKQRELQERMSR-----VQGRSGLSMSAARSEHEVSE 374
Db 16 QVNDVYAKLPYNPISFP-IDKSTDERELEESRWSGCVADGDLMLFLRAARMAAPQG 74
Qy 375 IIDGLSEQENI-----EKQMRQLAVIPPLMLDADQQRIFKINMGLMADPMKYKDRQVM 429
Db 75 MCDGLEDGGLAASRDDTTINALVDHSDGYPGKA-----LQALVKCPVSKGIDKK-- 126
Qy 430 NMWSEQEKETPREKFMQHPKFGILI-ASFLEKTVACVLVYIYTKQENYKSLVRYSR 488
Db 127 --WTEDETKKFIKGLRFGKGNFFRIHKDLLPHKDTPELVFYIYLVKKTTPGANN--NRPHR 182
Qy 489 RRGKSQQQQQQQQQQQQQQQQQQPPRSSQBEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 548
Db 183 RRRGSLRRNRVTRANNSNWTP-----PKKEDTPEPOT----- 216
Qy 549 KTDGTSGEDNDEKAVASKGRKTANSQGRKRGRITRSMANFANSEEAITPQQAELASME 608
Db 217 -----ATTATAATAASE-----TASRSSPAVSKE----- 241
Qy 609 LNESSRWTEEB-METAKKGLLEHGRNWSAIARMWGKTVSOCKNFYNYKRONLDEILQ 667
Db 242 --ENSSLTDDASECDSLSLTHKRDESP--SRM-----RTRN----- 275
```

```
Qy 668 QHKLMEKERNARRKKKAPAAASEEAAFPVVEDEEMEASGVSGNEEMVEEAEALHAS 727
Db 276 -----KQNNNSSTSGNNTAGCGGNATSISSGSTGGGAAGCNSSSKQDSANAV-AN 327
Qy 728 GNEVPRG-----ECSGPATVNNSSDTEIPSHHTAAKDTGONGPKPPATLADGPPGPP 783
Db 328 GKRPRGSETPDVSGGASVDS-----PKTPTTAKVAESSANKRK-----GKQDTP 372
Qy 784 TPRRTSRAPLEPTPASEATCAPTPPPAPPSP-----SAPPPVVPK-ESEKEETAAP 835
Db 373 NKKRTQESNEPS-AHEENAIKKRKRKPDSPVSMNSDSRPSDVLDDGENTTTTAA- 430
Qy 836 PVEEGEQKPPAAEELAVDTGKAEPEVKSECTEAEAGPAKGDAEA-AEATAEGALKA 894
Db 431 -----EQSTKDSKE-TVSCKEEREMVINDLEAKAE--KAIAEALAEADSKSAIKNM 481
Qy 895 KKEGSGRATTAKSSGAPQSDSS-----ATCSA--DEVD 927
Db 482 DEE-----TNIQAPSSADTSLVDGPNALPSPVAAPITMKVPTIATVEALNASVD 532
Qy 928 EAEGGDKNRLLSRPRLTPTGDPANASPOKPLDLKOLKORAAAIPIQVTKVHEPRE 987
Db 533 RKEAIERMESCDSPMLKLATIKQEVSPQQQHMQQSQ-----QQMQQ 578
Qy 988 DAAPTK-PAPPAPPPQN-----LOPESDAPQPGSSPRGKSRSPAPPADKFAFAEA 1039
Db 579 QLAPVGIPQPPSCPSPSVYIKPEMEDSMDCATCNQNSNEPDLKVKIEIKNEDALKHSA 638
Qy 1040 QKLPDPPCWTSGLPFPVPPREVIKASHPADPSAFSAPPG--HPLPLGLHDTARPVLP 1097
Db 639 GGLPPSGPC-----APPSALHPL----- 656
Qy 1098 RPPTISNPPPLISSAKHPSVLEROIGALSQMSVOLHVYPSE-HAKAPVGPVTMGLPLPM 1156
Db 657 -----SGAP--VESGQPLHLQ-----HMPHQVTTQPPPGVYLIDG----- 690
Qy 1157 DPKKLAPFSGVKQQLSPRGQAGPPESGLGVPTAQEASVLRGTALGSGVGGSTTKGIPST- 1215
Db 691 -----QLKYSGSGGVPP-----QPQLHSDAAGGVSGA--PPGAPITP 727
Qy 1216 -RVPDSAITVRSITHTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEYK 1274
Db 728 QKYPPEMEMKF-----APQDLKY-----PPPPPLD----- 752
Qy 1275 GHVLSYEGMSVTQCKEDGRSSSGPPHETAPKRTYDM-----MEGRVGRAISSASIEG 1329
Db 753 --ALKYSQEMQAAAA-----AAAAKYDMKYMMEQOQKYNVELSAH--- 793
Qy 1330 LMGRAIPPER-----HSPHLKEQHIIIRGSIQTGIPRSYVEAQEDYLR 1372
Db 794 -----QPPSKPGYQDSLKIPIKPGFGLPHNV-----GSP--LDAAHKY-- 831
Qy 1373 REAKLLKREGTPPPPPSRDLTEAYKTQALGPLKLPKPAEGLVATVKEAGRSIHIPREE 1432
Db 832 -----GPPPTSQESQQQFQ-----PPAHQVPPGATPPGIA--MPKPH 868
Qy 1433 LRHTPELPLARP-----LKEG---SITQGTPLKYD-----TGASTTGSKKHDVRS 1476
Db 869 YQHDVQTPPLGRPEFTGLMLKYGDPLAAKYGGPQDLKYPMPVPSQAGPADVKPYGGNL 928
Qy 1477 IGSPGRTFPVPHVDVNMADARALERACVEESLKSRRPTASSSGSGSIARGAPVIVPELKP 1536
Db 929 IKSPYGFPPESPIDA-----SARSTFG--QDSQGSNSNSQPSMPPQPOQ 972
Qy 1537 RQSLTYTHDGAPFAGHLPRG-----SPVTMREPT----- 1566
Db 973 FQSPHSPHMPSPAGGGLPGMHQPNLIHGPPGAAGSGGPPQPPPTSLHQPTPSAGP 1032
Qy 1567 PRLOEGSLSSKASQDKLSTPREIAKSPH--STVPEHHHPH-ISPVEHLLRGVSGVDL 1623
Db 1033 PSLOHGLHPGHQHSQALSASSIPSSSIGIPTLTJSTMAPSHMHLPHAH--LQGL----- 1086
```



```
QY 1765 -POPFSSRRHSSPLSPG-----PTHL-----TKPTTS 1792
Db 1234 HPGHLLSHSTAGLPGCGGPIALLAGPGLGIPESALSRRTPSHLPHSHASSAPLTAH 1293
QY 1793 SSERDRDRDRDREREKSLTSTTVEHAPWR--PGTEQSSGSSGSG-----CGGG 1846
Db 1294 SVA-----SMTSTSMSTLTSTVPSSAFSRASPSVQISSGGSGSGSGVPGG 1341
QY 1847 ---SSSRPASHSHAH-----QSPISPTQDALQORPSVLHNTGKGIITAV 1890
Db 1342 MPNSAAAAAHHRAASPASSVSSLSRQPLHPVQSPLSHPS-----SSALSAAAAV 1398
QY 1891 -EPKPTVLRSTSSVFRPAATFPATHCPL-GCTLDGVYP-----TLMEPVLLP 1939
Db 1399 AERDRHALMQOS-----PHMTFPVVSNAASLASPLSKWYAPQOGQGLGTSPPHLP 1452
QY 1940 KEAPRVAR--PERPRADTGHAFAPKAPARGLEPASPSSKSEPRPLVPVPSGHATARTP 1998
Db 1453 GASPPVIRHPOMPL-----PULIAPGGGIPQIGVHP 1484
QY 1999 AKNLAPHASDPAPPASADPHREKTQSKPFSIQELELSLGYHSSYSPEGVVPSP 2058
Db 1485 GQSPYH-----PLLHPSVFVSPHH-----PFN-----SPYGYAPYGP 1518
QY 2059 VSSPSLTHDKGLPKHLELDKSHLEGLRKPQGPVKLGG-EAAHLPHLRPLP-----ES 2112
Db 1519 -----GFPAYMKP-----PPQQLDPAVAAWAAHAGLQGPFPQOMRQD 1557
QY 2113 QPSSSPLLQTAGVKGHQVVTLAQHISEVITQY-----TRHHPOOLSAPLPAP 2162
Db 1558 EQNAAAAAAQAEEKQHAQAAAAAQAQKAPQOQPGGMPNKPPTPKTPQGGGMP-- 1615
QY 2163 LYSFPGASCPLDLRRPSSDL-----YLPPPDHG--APARGSPHS-----EGKR 2205
Db 1616 ---PGMGGTPPTGLPGAYPGSHMFCYPQGPFGHGFAPQDQPHGLKFTSHMDALRA 1671
QY 2206 SPEPKTSVLGGGEGDIEPV-----SP-----PEGW-----TEPGHRSR 2240
Db 1672 HAHSANSAGMGHGHTEPLDIDIEPEPEIPSTTHNIPRGSPSEAKPDDECHRSOSA 1731
QY 2241 VYPLYRDGEOTEPFRMGSKSPGNTSOPPAFFSKLTESNMVSKKQKQKINKLNTNRN 2300
Db 1732 IFVRHIDRGDYNSTR-----TDLIFKPVADSKLA---RKREERDRKLAERE 1777
QY 2301 EPEYNISQPGTEIFNMPAITGTGLMTRSQAVQSHASTNMGLEAIIRKALMGKYDQWES 2360
Db 1778 RROQOQOQ-----QOQOQOQAAAAQAQAQAQA-----ELK 1811
QY 2361 PPLSAN-----AFNPL-----NASASL-----2377
Db 1812 PPYADTPALRQLSEYARPHVAFSPVEQMVVYHHPGMPYRERELEIKNAQAAAAQSRL 1871
QY 2378 -----PAAMPITAADGRSDHLLTSPGGGKAKVSGRPSRKAAPSAPGLASG 2424
Db 1872 DPHWMEYYRGIHPSQFLYANPAISOEMERLERLGIPPPHVGLDPGEHVMRMPQPEAGF 1931
QY 2425 DRPPSV 2430
Db 1932 QLPPNV 1937

RESULT 28
O95875 PRELIMINARY; PRT; 2157 AA.
AC O95875;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE BAT2.
OS Homo sapiens (Human),
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```

[1]
RN SEQUENCE FROM N.A.
RP Rowen L., Madan A., Qin S., Shaffer T., James R., Ratcliffe A.,
RA Abbasi N., Dickhoff R., Loretz C., Madan A., Dors M., Young J.,
RA Lasky S., Hood L.;
RT "Sequence of the human major histocompatibility complex class III
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF129756; AAD18086.1; -
SQ SEQUENCE 2157 AA; 228669 MW; 21B817F5B699B0DB CRC64;

Query Match. 4.1%; Score 546.5; DB 4; Length 2157;
Best Local Similarity 20.4%; Pred. No. 8.9e-18;
Matches 514; Conservative 207; Mismatches 871; Indels 927; Gaps 114;

QY 474 KKNENYKSLVRSYRRRGKSKQQQQQQQQQ---QQQQQQQPMRPSSEKDEKEKEAE 530
Db 70 KGNDPNVSLPKDGTGWSAQEQSDPKSDASTAQPPESQPLPAS-----QTPAS 119
QY 531 KEEKPEVENDKEDLLKEKTDGSDNDEKAVASKRKTANSQGRKRTANSQGRKRTITRSMANE- 589
Db 120 NQKRPAAAPENTPLV-----PSGV---KSWAASVTHGAHGDGGRASSLSRFSREEF 170
QY 590 -----ANSEEAITPOQSAELA-----SMEINSSRWTE-----EMETAKGLLSEHR 632
Db 171 PTLQAAGDQKAAKAKERESABQSGPGSLRPQNSTTWRDGGGRGPDELE-GPDSKLHHGH 229
QY 633 NWSA-----636
Db 230 DPRGLQSPGPPFPYRGMPPFMYPPYLPFPFPPYQPGPYPTDPGSRFRVAVAGR 289
QY 637 -----IARVSGSKTVSQCKNF--YFNYKKRONLD-----EILQOHKLMEKERNARR 681
Db 290 GSGPPMELVFPVGRPSILKEDNLKEFDLQDQNDGWAHAEEVDYTEKLKFSDEEDGRD 349
QY 682 KKKK-----APAAASEEAAFPVVEDEMEHAGSVSCNEEEMVVEAEALHASNEVPRG 734
Db 350 SDEGAEGRHDSOSAGEER--PP-----EADGKGN-----SPNSBPPT 387
QY 735 ECGSPATVNNSSDTE-----SIPSPHTAAKDTGONGKPKPATLGDGPPPGPTTP 786
Db 388 PXTAWAETSPPETEPGPPAPKPLPPPHRGAGNAGWPGPDYP-----DRGPPCKPAPE 443
QY 787 -----RRTSRAPIE-----PTPASE 801
Db 444 DEDEAWQRKQSSSEISLAVARRRRREERREERAAEAKELRLDEKFGADPKR 503
QY 802 ATGAPTPPPAPPSAPPPVVPVVEKEEETAAAPPVEEGEOKPPAAEELAVDTGKAEEP 861
Db 504 LKAPPAAPPAAPSTPAPPVAPVKE-----LPAPPA-----PP-----P 536
QY 862 VKSECTEEAEGBGPAKDAEAAEATAEALKAEEKGSGRATTAKSGAPQDSDSATC 921
Db 537 ASATPEKEPEEPAQAAPPAQSTTTP-----GVAAAPTIVSGGSTSTSSGSF 584
QY 922 SADEVD-----EAEGGDKNRLLSPRSLTTPGDRANA-----SPQKPLD 962
Db 585 EASVPVQLPSKEGPEPEPEEVPPTTPPVKVEKGDIGTRQPPQSGLYPKYQKSLP 644
QY 963 LKQLKQRAAIPPIQVTKVHEPPREDAAPTKAPAPPPPPONLOPESDAPQPGSSPRGK 1022
Db 645 PRFORQOQOQLLKQOQOQHQWQHQOQSGSAPPTVPVPPQPVTLL-GAVPAPOAP-----696
QY 1023 SRSPAPPADKEAFAAEAQKLPDPPCWTSGLPFPVPPREVIKASPHAPDPSAFYAPGPH 1082
Db 697 -----PPPPKALY-----PG-----ALGRPPMPMPNPF-----DPRWMIIPFYVD 731
QY 1083 PLPLGLHDTARPVLPRPPTTISNPPLISSAKHPSVLRQIGAIQSGMSVLQHPYSEHAK 1142
Db 732 P-----RLQGRFPFLDFYPPGV-----HPSGLVPRERSDSGGSSE---PDRHAP 774
QY 1143 APVCGVTWGLPLPMDPKLIA-----PFGVKQEQ---LSPRCQAGPPESLGYPTAQEASVL 1195
Db 1143
```

Db 775 AMLR--ERGTP-PVDP-KLAWGVDTATPAEPRPLTSLPRLQAEDDDKGM----- 821

Qy 1196 RTALGVSFGSITKIGFSTRVPSDAITRGSITHGTADVLTKGTITRI-IGSDSPSR 1254

Db 822 -----RSETPPVP--PPYLASYPGFENGAGP-----PISRPLEPGPRP 863

Qy 1255 LDRGREDSLPKGHVIEGKGVHLSYEGMSVTQCSKEDGRSSGPPHETAAPKTYDMM 1314

Db 864 L-----FWPPG-----SDEVAKITQTPPKPKPPKEETAQUT 894

Qy 1315 EGRVGAISSAIEGLMGRAIPPERHSPHLKQHHIRGSITOGTIPRSVVEAQEDYLRE 1374

Db 895 GPEAGR--KPARGVSGCGGPPPPRRESTRWGRFGSSRRGIPP----- 939

Qy 1375 AKLKAREGTPP-----PPPSRDLTEAVKTOALGLKLKPAHEGLVATVKEAGRSIH 1426

Db 940 -----EPCAPPRRAGPIKKPPPT-----KVELPP--KPLAQ-----D 973

Qy 1427 EIPREELRHTPLAPRLK-----EGSITOGTPLKYDTGASTTG----- 1467

Db 974 ETPK-----PKPDPLKITKGLGPKETPPGNLSAPRLRRDYSYERVGPTSCR 1024

Qy 1468 -----SKXHDVRLIGSGRTPPPVHPLDVMADARALERACVEESLKSRCGTASSG 1519

Db 1025 GRGGEYFARGRGFTYGGRR-----GARSREFRSYRE-FRGDDGRGGGTG 1071

Qy 1520 GSIARGAPVIVPELGPQRSPLTYEDHGAPFAGHLPRG--SPVTMRE-----PTPRLQ 1570

Db 1072 G-----PNHPP-----APRGRTASETSEGESEYEELPKRRQ 1103

Qy 1571 EGSLSSSKASQ-----DRKLTSRPE--IAKSPHSVTPEHHPIPSVYEHLLRGVSGVDL 1623

Db 1104 RGSETGSETSDSLAPSDEAPTKEGTLTQVPLAPPPGAPPPAPARFTARG----- 1157

Qy 1624 YRSHIPLAFDTSIPR-----GTPLDAAYLPL--RHLPN--PTYPHLYPP----- 1667

Db 1158 -----GRVTPRGVPSRRGRGGRPPOVPCGWSPPAKSLAPKPPPTGP--LPSPKELK 1210

Qy 1668 -YLIRGVPDTAALENRTIINDYITSQMHNTATAMAQRAQMLRGLSPRESSLAL--N 1723

Db 1211 EKLIPGLPSVARGSGNSGVNMGEDGERPRRRHGRAQQQDK-----PPFRRLKQREN 1266

Qy 1724 YAAGRGIIDUSQVPHLPVLVPTTGTATAMDRLAYLPTAPQPPSSRHSSPSLPGOPT 1783

Db 1267 AARGSEKPSLT--LPASAP-----APEEA--LTTVTAPAPERRAAKSPDLNQNSD 1315

Qy 1784 HLTKPTTSS-----SERERDRRE-----RD 1805

Db 1316 QANBEWETASESDFTSERRGDKAPVPLVLTAKVTPGGGGGAVGIGSAMSRLDLSQ 1375

Qy 1806 RDREREKSIILTTTVEHAPTWRPQTEQSSGSSSGSGGGG----- 1846

Db 1376 RAKDLKRESFSSQRPQMERQNRFPQCGKAGSSSSSGGGGPGGRTGPGRGDKRSWPS 1435

Qy 1847 --SSSRPASHAHQHSPISPRTOALQORPSVLHNTGMKGIIITAVPSPKTVLRSTST 1904

Db 1436 PKNRSRPEERPPGLPLPPLPPSSSAVRLDQVIH-----SNPAGIQALAQ 1482

Qy 1905 SPVTPAATFPATHCPLGGLTLDGVYPTLMEVLLPKAPRVARPERPADTGHFLAKPP 1964

Db 1483 LSSRQGVTPAGGH-----PR-HKPGPPQAPQGPS--PRPP 1515

Qy 1965 AR-----SGLEPASGPSKSEPRLPVPPVSG--HATARTPAK----- 2000

Db 1516 TRYEPQVNSGL-----SDPHEEPGPWVGVTGTPRDSAGVSPPPKRRERPPKPELLQ 1572

Qy 2001 --NLAPHAS-----PDPPAPPASAD-----PH-----REKTSKPSIQELE--- 2037

Db 1573 EESLPPPHSGFLGSKDEGPGQAESRDTGTEALTPHWNRLHTATRSKSYRPSMPEWM 1632

Qy 2038 -----LRSLGHGSSYSPEGVE-----PVSVPSSSLTHDKGL-----PKH-- 2073

Db 1633 EPLSPFEDVAGTEMSQSDSGVLSGDSQVSSGPGCSQRSSP-----DGLKGAAEGPPKRRP 1688

Qy 2074 ---LEELDKSHLEGELAPKQPGPVKLGGEAAHPLHLPLPESQSSSPSLQLQTAPGVKXH 2129

Db 1689 GSSPLNAVPCGPGSEPPRPPAPHDGRKELPREQLP-----PGPIGT 1735

Qy 2130 QRVVTLAQHISEVITQDYTRHHPOOLSAPLPAPLYFFGASCPVLDLRRPPSDL----- 2183

Db 1736 ER---SQH-----TDRGTEPGPIRPS-HRPGPPVQF-GTSDKSDLRLLVVGDSLKAKE 1784

Qy 2184 -----YLPDPDHGAPARG-----SPHSEGGKESPE-----PN 2210

Db 1785 LTASVTEAIPVSRDWELLPSAAASAEOSKNLDSGHCVPPESSSQRLYPEVIFYGSAGPS 1844

Qy 2211 KTSVLGGGED-----GIEPVPSP-----PEGMTEPGHRSASV-YPLLYRDG 2249

Db 1845 SSQISGGAMDQLHPNSGGFRPGTSLHPYRSQPLYLPPGPAPPSALLSGVALKQFLDF 1904

Qy 2250 EQTPSRMGSKSPGNTSQPPAFF-----SKUTESNAMVYSKKQKQINKKLNTNRNE 2301

Db 1905 STMQATELGKLPAGGVLYPPSPFLYSFAFCPSPLPDTSLQVRQ----- 1948

Qy 2302 PEYNISOPGTEIFNMPAITG--TGLMTYRSQAOEHASTNMGLEAI IRKALMGKYDQWEE 2359

Db 1949 ---DLPSP-SDFYSTPLQPGQSGFLPSGAPAQ-----MLLPWVDSQLPVVNFGLSP 1998

Qy 2360 SPPLSANAFNPLNASASLPAAMPITAAADGRSDHTLTSPGCGKAKVSGRPSRSKAKSPAP 2419

Db 1999 APP---PAPPLSLLPVCPALQPPSLA-----VRPPAPATRVLPSPA- 2038

Qy 2420 GLASGRPPSVSVHSE-----GDCNRRTPLTNRVWEDRPSSAGS 2459

Db 2039 -----RPFPAALGRAELHPVELKPFQYQKLSNGLGPGSSRTPPTGRSFGSLNRLKA 2092

Qy 2460 TPFYNYPLI-MRLQAGVMASPPP-----PGLPAGSGP---LAGPHHAWDEEP 2502

Db 2093 TPSTYSGVFTQRVDLYQQASPPDALRWIPKWPRTGLPPREGPSRAAEFGSGDKPEP 2151

RESULT 29

Q9P3J0 PRELIMINARY; PRT; 2592 AA.

AC Q9P3J0;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Conserved hypothetical protein.

GN B7F21.040.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariales; Neurospora.

OC NCBI_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Aign V., Hoheisel J., Brandt P., Partmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.,

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -!- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.

DR EMBL: AL389901; CAB97476.2; -

DR GO: GO:0005634; C:nucleus; IEA.

DR GO: GO:0003677; F:DNA binding; IEA.

DR InterPro: IPR001005; Myb DNA binding.

DR Pfam: PF00249; myb DNA-binding; 2.

DR SMART: SM00717; SANT; 2.

DR PROSITE: PS50090; MYB_3; 1.

KW Hypothetical protein; DNA-binding; Nuclear protein.

SQ SEQUENCE 2592 AA; 288030 MW; 22CC6601A164A2D8 CRC64;

Query Match

4.1%; Score 543.5; DB 3; Length 2592;

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
Hypothetical protein KIAA1555 (Fragment).
KIAA1555.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
DR EMBL: AB046775; BAB13381.2; -;
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR002034; AIPM/Hcit synth.
DR InterPro: IPR007087; Znf C2H2.
DR Pfam: PF00096; zf-C2H2; 2.
DR SMART: SM00355; Znf C2H2; 5.
DR PROSITE: PS00815; AIPM_HOMOCIT SYNTH 1; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE: PS01517; ZINC_FINGER_C2H2_2; 5.
DR Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 2414 AA; 260294 MW; 05215E33CCEE2B65 CRC64;
Query Match 4.1%; Score 536.5; DB 4; Length 2414;
Best Local Similarity 19.6%; Pred. No. 3.1e-17;
Matches 546; Conservative 316; Mismatches 945; Indels 973; Gaps 135;
QY 7 LVAQTWATEPRYPHSLSYVQIARTHTDVCILLEYQHHSDYASHLSPGSIIOQRRRP 66
DB 123 LEGSTWLVPMRPGSGFVAP-----GL--HPQSLLPESHAS-----IIPEDLP 167
QY 67 SLASEFOFGNRSOELHRLPESHVLPGLGKSEMEFTESKRPR-----LELLPDPLLR 120
DB 168 GVPKVFVP---RPSQVSLKPTEEAHKE-----RKQPKGVICQYCSRPAKP 213
QY 121 SPLIA-----TGO-----PAGSBDLTKD-----RSLTGKLEVPSPPPHTDP 158
DB 214 SVLQKHIRSTGERPYPCGPGCFGFKTKNLYKHKRSHAHRIKAGLASGMGEMYPH-GL 272
QY 159 ELELVPPRLSKEELIQNMDRVDRITVWEQOIKLKKQQOLEEAAKPEPEKVPSPFP 218
DB 273 EMERIPGEEFEEP-----TEGSTDESETSATSGHPAELS-----PRPKQPLSSG 319
QY 219 IES--KHSLSVQIITYDENRKAEEAARILEGLGQPVELPLVNOPSDTRYHENIKINQAM 276
DB 320 LYSGSGSHSSHERCSLSQSSTAQSLEDDPPFPFVPSSEHPLSHKPEDTHTIKQJLAULSE 379
QY 277 RKLLI-----LYPKRNHARKQWKQKFCORYDQMLEALEKKVVERIENNP 320
DB 380 RKXVIDQAFLLSPGSKGSTSGYFSRSEAEQ-----VSP 416
QY 321 RRAKESKVEYVEKQPETRKQRELOERMQSRVQRGSGLSMSAAR-----SEHEVS- 373
DB 417 NTNAK-----SYAEIIFGKC-----GRIGQRTAMLTATSTQPLPLSTEDKPSL 460
QY 374 -----EIIDGLSEQENLEKQMLQVAPPMLYDADQRIKFINNGLMADPMKYKD 425
DB 461 VPLSVPTQVIEHITKILITNE-----AVDVTSEIDSVKPRRSLRSMESP-KESLY 514
QY 426 RQVMNMWSEQKETFREKFMQKNFGLIASFLERKTV--AECVL-----YYLTIK 474
DB 515 REPLSSHSEKTKPEQSLLSLQHPSTAPPVPLLRSMSPSNACTIPTPHPRFGSYSD 574
QY 475 KKNYKSLVRSY-----RRGKSGQQQQQQQQQQQQQQMPRSGEEKDEKEKEKEAE 530
DB 575 HTDTSALSSSSHVFTSHPRMLKRAQPAIELPLGGEYSSEEPFGSS-----KDTASKPSD-- 628

Qy	531	KEEKPEVENDKEDLLKEKTDGTDGONDEKEAVASKGRKTANSQGRKG----	RI	TRS	585
Db	629	-----EVE-PKESELTKKT-----	KG	KL	KT-----
Qy	586	MANFANGSEALPTOOSAEELASMEINBSRWTEEMETAKGLLBHGRWNWAIARMVGSKT	645		
Db	660	RYKRDNYEAHKYCYSELQIAKISAGHTTSP-----AEKQIEH-EWVSQ-----	706		
Qy	646	VSQCKNFYNYKQRNLDEILQOHLKWEKERNARRKKKAPAAAABEAPPPVVEDEEM	705		
Db	707	-----MMHYKLTGTTLETLPLRKKRKESLGDEEE-----PPAFESTKS	744		
Qy	706	EASGVSGNEEMVEAEALHASGNVEPRGECSPATVNNSSDTEISIPSPHTEAAKDTQON	765		
Db	745	QFGS-----PGPSDAARNLPLESTKSP-AEPSKSV-----	773		
Qy	766	GPKPPATILGADPPPGPPTP-----PRTSR-----APIEPTPASEATGA-----	805		
Db	774	-----PSLEGPTFGFQRTPKPGSGSESGKERRTTKEISVIOHTSFEKSDSLEQPSGLEG	829		
Qy	806	-----PTPPPAP-----PSPSAPPVVPVPEKEEBEATAAAPVVECEE	842		
Db	830	EDKPLAQFPSPPPAPHGSAHSLQPLVRQPNIOVPELLVTEEDRDPDTEPEPPKPEK	889		
Qy	843	QKPPAELAVDTKASBEPVKSECTEE--ABEGPAKGKO-----ABAAEAABGALKAEKKE	897		
Db	890	-----TEEFQWORSQTLAQPAEKLPPKKRLRLAEWAQSSGESSPESSVP-	936		
Qy	898	GGSGRATTAKSGAPQSDSATCSADEVDEAEGDKNRLSP-----SLTTTGDP-	951		
Db	937	--LRSFSQSNVSLSGSSRSASPERDDHGAEPSPSSMRPKPLGTHMLTVPSHHPHA	994		
Qy	952	---RANASPOKP-----LDLQKLQRAAAIPIQVTKVHEPPREDDAAPTKEPAPP--	1000		
Db	995	REMERASSEQPNVSHSAHMTETRSFDYGSLSLT-----GPSAPVAPPARV	1044		
Qy	1001	PPQNLQ-----PESD--AP--QOPGSSPRGKRSRSPAPPADKFAFAEAQKL	1042		
Db	1045	APPERRKCFVLQRASLRPPESELEVAAPKGRQSEBEPQPSKPSAKSLSQISAAATSH	1104		
Qy	1043	PGDPPCWCTGLPPP-----VPPREVKA-----SPHADP-SAFSY-----	1077		
Db	1105	GGPPGGKGGQDRPLGPTVYTEALQVFIHPVNAQTLHKKPYLPPVSVLSFQHLVOHE	1164		
Qy	1078	-----APPGHP-LPLGLHDTARPVLPRPTTISNP-----PP	1107		
Db	1165	PGQSPPEFSTQAMSSLLSSPYMPLLPPLFQA--PPLPQPTVLHPGQLHLPQMLHPHA	1222		
Qy	1108	LISAKHPSVLERQI--GAISQGM-----SVQLHVPVSEHAKAPVGPVTM	1150		
Db	1223	NIPRQPPSFLMPWPTTSSALSQGFPLQSQFALQPGDVESHLPOKTLAPLATGSA	1282		
Qy	1151	GLPLPMDPKLAPFSGVKQE-QLSPRQAGPPESLGVPTAQEASVGTALGSPVGSIT	1209		
Db	1283	G-----LSPESTEYSSDIRLPP---VAPPASSAPTAPP-----LALPACPDMTVS	1325		
Qy	1210	KGIP-----STRVPDSOAIITYRG-----SITHGTADVLY-----KGITRIIGBD-	1250		
Db	1326	LVVPVRVQTNMPSYGSAMYTTLSQLVTQSGSSATVALPKFEBPPPSKG--TTVCGADVH	1383		
Qy	1251	-----SPSRLDRGHEDSLPKGH-----VIYEGKKGHVLSVEGMSVTQCSKEDGRSSGPPH	1302		
Db	1384	EVGFGPGLSEBQSRAPFTTYLVRVPVTLPERKGTLSSESILSL-----EGSSSTAGGSK	1438		
Qy	1303	ETAAPKRTYDM-----MEGRVGRAISSASIEGLMGRAPP-----ERHSPHLK	1346		
Db	1439	RVLSPAGSLELTMTQOQKRVKEEASKADEKL--ELVKPCSVVLTTSTDGKREKSHL-	1495		
Qy	1347	EQHIIRGSITQGIIPRSVVEAQEDYLRBKAILLKREGTTP-----PPPPGRDL-	1393		
Db	1496	-----GNQOG-----RRELEMLSSLDSPDQTKIEIPLPHPALSHGTAPG	1536		

QY	1394	TEAYK	---TQALGPBKUKPAHEGLVATVKEAGRSIHEIPRELRHTTPELP	---RPLK	1444	
Db	1537	SEALKPEYPSG	---KPHRGLTPL	---SVKEDSKQDPLPSLAPFSSLSPLS	1583	
QY	1448	EGSITQGTPLKYDTGASTTCSKKHDVRSLLIGSPGRFPFVVH	---PLDVM	1493		
Db	1584	E---TSGRPAKSQEG	---TDSKK	---VLOFFPSLUHTTNNVSWCVLNVIKNHHIQ	1627	
QY	1494	ADARALERA	---CVEESLKRPGTAGSSGGSIARGAPVIVPELCKPQRSPLTYEDHGAPF	1550		
Db	1628	HABRRSVYAGWICISLYPNPLPGVSTKAAUSLLR	---SKQVSKETYTMATAP	1677		
QY	1551	AGHLRGSPVTMBREPTPRLOE	---GSLSSSKASQDRKLTSTPREIAKSPHSTVP	1604		
Db	1678	--HPEAGRLVPSSSRKPRMTEVHLPSLVSPGKDLARVEKEERRGEPEADAPASQORGE	1735			
QY	1605	PHPISPYEHLLRGVSGVDLVRSHIPLAF	---DPTISIRGPIPLDA	1645		
Db	1736	PARIKIPE	---GG	---YKSNEEYVYVVRGRGKYVCBECGIRCKKPSMLKKHIRTHT	1786	
QY	1646	AAAYVILPRHL	---APNPTYPHLYPPLIRGVPDTAALENRQTIINDYITTSQOMHNTA	1700		
Db	1787	DVRPYCVCKHCHFAFKTKGNLTKEWKSXAKSKKQCTGVLE	---ELEAEG	1833		
QY	1701	TAMAQRADMLRGLSPRESSIALNYAAGPRGI	IIDLSQVPHLPVLVPPPTGPTATAMDRLAY	1760		
Db	1834	TS	---DDLQDSEGESEAV	1851		
QY	1761	LPTAPQFSSRHSSPLSPGCPHTLTKPTTTSSERRDRDRDRDRREKREKILTSITT	1820			
Db	1852	---BEHOFSDL	---EDSDSDSLDEDEDEESQDELGRPSS	1888		
QY	1821	---VEHAPIWRP	---GTEOSSGSSGS	---SGGGGSSSRPA	1852	
Db	1889	EADPPGPPHALRADSSPILGQPPDAPAGTEATRGSSVSEAEELTASSCSMSQNPGL	1948			
QY	1853	---SHSHAHQHSPTSPR	---TQDALOORPSVLHNTGMKGIIITA	1889		
Db	1949	PWLGPAPLGSVEKDTGSALSYKVPSPRRPWSPSKEAGSRPPLAKHLSLTKNDS	2001			
QY	1890	VEGSKPTVLASTSTSSPV	---RPAATFPFATHCPLG	1922		
Db	2002	SPQRCSFAPREPOASAPFPGLHVDQGRGMGLPCGSPRLQLSLTLCPLGRELAPRAHV	2060			
QY	1923	---GTLGDVYPTLMEPVLLEKPAE	---RVAPERPRADTGHAFKAPPARSGLE	1970		
Db	2061	LSKLEGTID	---PG	---LPRVSPTRRWSPGQAESPPR	2101	
QY	1971	PASSPFSKGSE	---PRLVPPVS	---GHATIAITPAKNLAPHHA	SPDPPAPPASADSP	2021
Db	2102	GPGPSAGEHCPGLIAPRVLFPAPLPHKLLSRSPETCASPWKAESRSPSCSPGPA	2158			
QY	2022	HREKTQSKPFSIOLELRSLRGYHGSSYSPEGEVPSVPSPSLTHDKGLPKHLELDKSH	2081			
Db	2159	--HPLSRPFS	---ALHDFHGHTLARTENI	---FSH	---LPLHSQHLTRA	2198
QY	2082	LEGELRPKQGP	---VKLGGEAAHLPHLRPLSPQSSPLLQTAPE	---GVXGHQRVVT	2134	
Db	2199	---PCPLPIG	---IQMVQARPCAHPTLLPGTAAVWSFGSGGSDLT	2241		
QY	2135	LAOHISEVITQDYTRHHFQQLSAPLAPLXSPFGASCFLVLDLRPPPSDLYLPPPDH	2190			
Db	2242	GARE	---AQERGRWSPTESS	---ASVSPVAKVSKFTLSSELEGGDYPER	2286	
QY	2191	---CAPARGSPHSEGGKRSPEPNKTSVLGGEGDIEPVSPEGWTPEGHSRSAYVP	2243			
Db	2287	ERTGGGCRPPDWTPHGTGAPAEPTTHS	---PCTEPDTPRPPQGR	2330		
QY	2244	LLYDGEQTEPSRMGSKSPGN	---TSOPPAFFSK	---LTESNAAWVKSKKOEINKKLNT	2296	
Db	2331	---RAAQSWSPRLSPRANPEPESATPDLDRSSVVCGLAFA	---SARFPARTNLUSGERT	2386		
QY	2297	H	---NRNEPEYNISQ	2309		

Db	2387	QDSPKPSGSGEPRAHPHQ	2406
		:	
RESULT 31			
Q96QC6		PRELIMINARY;	PRT; 2157 AA.
ID	Q96QC6		
AC	Q96QC6;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DE	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	BAT2 protein.		
GN	BAT2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RP	Shina S., Tamiya G., Oka A., Inoko H.;		
RT	"Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region.";		
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AP000505; BAB63391.1; -		
SQ	SEQUENCE 2157 AA; 228958 MW; C500E03BE15382C9 CRC64;		
	Query Match	4.0%;	Score 531.5; DB 4; Length 2157;
	Best Local Similarity	20.3%;	Pred.No. 4.7e-17;
	Matches 510;	Conservative 208;	Mismatches 870; Indels 927; Gaps 113;
Qy	474	KQENYKSLVRRYRRGKSGQQQQQQQQQ---QQQQQQQMPRSSGEEKDEKEKEAE	530
Db	70	KGNDPNVSLPKDGTGWASQEQSDPKSDASTAQPPSQPLPAS-----QTPAS	119
Qy	531	KEEKPVEVNDKEDLLKEKTDITDSGEDNDEKAVASKGRKTANSQGRKGRITRSMANE-	589
Db	120	NQPKRPAPAPENTPLV-----PSGV---KSWAQASVTHGAHGDGGKASLLGRFSREEF	170
Qy	590	-----ANSPEATTPQQAELA-----SMELNESSRWTE-----ERMETAKKGLLEHGR	632
Db	171	PTLQAQGDQKAAKERESASQSGPGPSLRPQNSITWRDGGRCFDPLE-CPDSKLHGH	229
Qy	633	NWSA-----	636
Db	230	DPRGGLQSPGPQPPPYRGMPPMYPPYLPFPYPYQGPQYRYPTPDGPSRFPVAGPR	289
Qy	637	-----IARVSGKTVSQCKNF-YFNKKRONLD-----ELQHKLKMEKERNARR	681
Db	290	GSGPPMLRVFVGRPSILKEDNLKEFQLDQENDDGWAGAHEEVYDYTEKLFKFSDEEDRD	349
Qy	682	KKKK-----APAAASEEAAFPVVEDEMEWASGVSGNEEMVEEAEALHASNEVPRG	734
Db	350	SDEGASGRHDSQASGEER--PP-----EADGKKGK-----SPNSEPPT	387
Qy	735	ECSGPATVNNSSDTE-----SIPSPHTEAAKDTGONGKPKPATLGAOGPPPGPTTPP	786
Db	388	PKTAWAETSRPPEPEGPAPKPLPPHGRGAGNWNPPGDYP---DRGGPPCKPPAPE	443
Qy	787	-----RTSPAPIE-----	801
Db	444	DEDEAWQRKQSSSEISLAVERRRRREERRMQERRAA CAEKUKRLDEKFGADPKR	503
Qy	802	ATGAPTTPPPAPSPAPVPVVKPEKEEETAAAPPVVEGEQEKPFAAEALAVDTGKAEEP	861
Db	504	LKAEPAPPAAPSTPAPPPAVPKE-----LPAPPA-----PP-----P	536
Qy	862	VKSECTEAEEGPAKGKDAEAAEATARGALKAEKKEGSGRATTAKSSGAPQDSDSATC	921
Db	537	ASAPTPKEPEEPAQAPPAQSTPTTP-----GVAAAPTIVSGGSGTSTSSGSF	584

Qy	922	SAEVD-----EAEQDKNRLLSPRSLLTPTGDDPANA-----SPQKPLD	962
Db	585	EASVPQLPSKEPPEEVPPTTTPVPKPEKGDIGITRQPPSQGLGPKYQKSLP	644
Qy	963	LKQLKORAAAIPIQVTKVHEPPREDAAPIKAPAPPPQNLOPESDAPOQCGSSPRGK	1022
Db	645	PRQROQOEQLLKQOQOQHQQOQAGPTTPVPPSPQVTL-GAVPAQAP-----	696
Qy	1023	SRSPAPPADKEAFAAEAKQLPGDPPCWTSGLPFPVPPREVIKASPHADPDSAFSAPPGH	1082
Db	697	-----PPPKALY-----PG-----ALGRPPMPMPNF-----DRNMWIPPYVD	731
Qy	1083	PLPLGLLHDTARVULPRPPTTSNPPPLISSAKHPSVLEROIGAISQGMVQLHVYPESHAK	1142
Db	732	P-----RLQGRPLDFYPPGV-----HPSGLVPRERSDSGSGSSSE-----PFDRHAP	774
Qy	1143	APVGPVTWGLPLMDPKKLA-----PFSGVKQEQ-----LSPRGQAGPPESLGPVTAQEASVL	1195
Db	775	AMLR--ERGTP-PVDP-KLAWGVDVFTATPAERPLTSLRQAADDKCM-----	821
Qy	1196	RGTALGSVPGGSITKGIPTRVPSDSAITYRGSITHTGTPADLVLYKGTITRI-IGEDSPSR	1254
Db	822	-----RSETPPVPP--PPPYLASYGPPENGAGCP-----PISFPFLEECPGRP	863
Qy	1255	LDRGREDLSPKGHVIEYEGKGHVLSYEGGMSVTCCKEDGRSSGGPPHETAAPKRTYDMM	1314
Db	864	L-----PWPG-----SDEVAKIQTPPKPEPPKEETAQLT	894
Qy	1315	EGRVGRAISSAIEGLMGRAIPIERHSPHILKEQHIRGSIITOGIPRSYVEAOBDYLRE	1374
Db	895	GPBAGR--KPARVGSGGQGPPIPPRRESRTETRWGPRPGSSRRGIIP-----	939
Qy	1375	AKLLKREGTPP-----PPPSRLDTAYKTAQLGLPLKKAHEGLVATVKEAGRSIH	1426
Db	940	-----EEGAPPRAGPIKKPPPT-----KVEELPP--KPLEQG-----D	973
Qy	1427	EIPRELRHTPELAPRPLK-----EGSITQGTPLKYDTGASTG-----	1467
Db	974	ETPK-----PPKDPDLKIKTGLGPKETPPNGNLSPAPRLRDSYERVGPTSCR	1024
Qy	1468	-----SKKHVRSILGSPGRTFPVPHPLDVMDADALERACYEESLSKRGTSASSG	1519
Db	1025	GRGGEYFARGRGFTYGGRR-----GARSREFRSYRE-FRGDDRGGGTG	1071
Qy	1520	GSIARGAPVIVELGKPRQSLTVEDHCAPFAGHLPRG--SPVTMBE-----PTPRLQ	1570
Db	1072	G-----PNHP-----APGRPTASRESEGSYEYEIPIKRRRQ	1103
Qy	1571	EGSLSSKASQ-----DRKLTSTPRE--IAKSPHSTVPEHHPII GYEHLLRGVSGVDL	1623
Db	1104	RGSETGSETHESDLAPSDKEAPTKEGTLTQVPLAPPCCAPPSPAPARTARG-----	1157
Qy	1624	YRSHIPLAPDPTSI PR-----GIPLDAAYILP--RHLPN--PTYPHLYPP-----	1667
Db	1158	-----GRVTFPRGVPSRRGGRRPPQVPCGSPSPAKSLAPKXPTGP--LPESKEPLK	1210
Qy	1668	-YLIRGYPDTAALENRQTIINDVITSQMHNTATAQAQADMLRGLSPRESSIAL--N	1723
Db	1211	EKLIPLGLSPVARGSGNGSVNGMEDGERPRRRHGEAQOQDK-----PPRFRLLKQEREN	1266
Qy	1724	YAAAPRGIIDLQVPHLVLVLPPTPGTPATA---MDRLAYLPTAPOFSSRHSSSPLSPG	1780
Db	1267	AARGSEG-----KPSLTLPASAPGPEEALTTVTVAAPRRAAAKSPDLSNQ	1312
Qy	1781	GPHTLKPTTTSS-----SERERDRRE-----	1803
Db	1313	NSDQANEWEWTASSESDFTSERRGDEKAPPVLLTPKAVGTPGGGGGAVPGISAMSRGD	1372
Qy	1804	-RDRDREREKSLTSTTTTVEHAPIWRPGTQSGSGSGSGSGSGGGGG-----	1846
Db	1373	LSORAKDLSKRSFSORPGMERQNRAPGPGCKAGSGSGSGSGGGGGGGGTGPRGDKRS	1432
Qy	1847	-----SSSRPASHSHAHQHSPISPRTQDALQORPSPVLHNTGMKGIITAVEPSKPTVLRST	1901

1433	Db	WPSPKNSRSEERPPGLPLPPPPSSAVFRLDQVIH-----SNPAGTQQA	1479
1902	Qy	STSSPVREPAATFPATHCPLGGTLGDGVYPTLMPEVLLPKEAPVAPRPRADTGHAFLA	1961
1480	Db	LAQUSSRQGSVTAPGGH-----PR-HKPGPPQAPQGPS--P	1512
1962	Qy	KPPAR-----SGLEPASSFKGSBPRPLVPPVSG---HATIAITPAK-----	2000
1513	Db	RPPTRYEPQVRNSGL---SSDPHEFEGPVMRVGGTPRDSAGVSFPPPKRERPRPKPE	1569
2001	Qy	-----NLAPHAS-----PDPAPPASASD-----PH-----REKTOSKPFISOELE	2037
1570	Db	LLQESLPPPHSSGFLGSKPEGPGQAESRDTGTEALTPHIWNRLLHATTSRKSVRSSME	1629
2038	Qy	-----LRSGLVHGSSYSPEGVE-----PVSVPSSPSLTHDKGL-----PK	2072
1630	Db	PWMEPLSPFFEDVAGTENSQSDGVDLSGDSQVSQSGPCSQRSSP---DGLKGAAEGPPK	1685
2073	Qy	H-----LEELDKSHLEGLRPKQPGVKVLGGEAHLPHLRPLPESQSPSSPLLQTAPGV	2126
1686	Db	RPQGSSPLNAVPCGPPGSPRRRPPAPHDGRKELPREQPLP-----PGP	1732
2127	Qy	KGORVVTLAQHISEVITQDYTRHHPOOLSAPLAPLYSPFGASCPLDLRRPPSDL---	2183
1733	Db	IGTER-----SOR-----TDRGTBGPRIIPS-HRPGPPVQF-GTSDKXDSLURLVVGSLKA	1781
2184	Qy	-----YLPDPDHGAPARG-----SPHSEGGKRSPE-----	2208
1782	Db	EKELTASVTETAI PVS RDWELLPSAASAE PQSKNLD SGCHVCFPSSSGQLYFEVYGSA	1841
2209	Qy	-----PNKTSYVLGGED-----GIEPVSP-----PEGMTBFGHSRSAV-YPLL Y	2246
1842	Db	GPSSQISGCMDSQLHPNSGGFRPGTPSLHPVRSQPLYLPCGPAPPSALLSGVALKGQF	1901
2247	Qy	RDEQTEPSRMSKSGCNTSQPPAF-----SKLTESNSAMVSKSQEINKKLNTHN	2298
1902	Db	LDFSTMQATELGLKAGGVLYPPPSFLYSPAFCSPLPDTLSLQVRQ-----	1948
2299	Qy	RNEPEYNISOPGTEIFNMPAITG--TCLMYTRSQAOEHASTWNGLEAIRKALMGKYDQ	2356
1949	Db	-----DLPSG-PDFTYPLQCGQGLFSGGAPAQO-----MLLPMVDSQPLVVNFGS	1995
2357	Qy	WEESPPLSANAPNLNASASLPAAMPTITADGRSDHTLTPSGCGGKAKVSGRPSSRKAKS	2416
1996	Db	LPAPP---PAPPPLSLLPVGPAIQPPLA-----VRPPAPATRVLP S	2036
2417	Qy	PAPGLASGDRPPSVSSHSE-----GDCNRRRTPLTNRWEDRPSS	2456
2037	Db	PA-----RPPFASLGRAELHPVELKPFQDYKQLSSNLGGPGSSRTPTTGRSFSGLNSR	2089
2457	Qy	AGSTPTPPYNPLI--MRLOQVMA SP-----PPPGLPAGSGPLAGHHAWDDEEP	2502
2090	Db	LKATPTSYSGVFTORVDLYVOOASPPDALARWIEKPKWERTGPPPPREGFSRR--ASEP	2143

RESULT 32
0911035

Q90Q35 ID 09U035 PRELIMINARY: PRT: 2752 AA.

Q9UQ35;
AC

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLRel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE RNA binding protein.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Ohtaki S., Umeki K., Sawada Y.;

RT "Homo sapiens mRNA for RNA binding protein, complete cds.";

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

Query Match	3.9%;	Score 521;	DB 11;	Length 2607;	
Best Local Similarity	19.1%;	Pred. No. 1.9e-16;			
Matches 550;	Conservative 325;	Mismatches 1091;	Indels 920;	Gaps 123;	
Qy	44	HHSRDYASHLSPG--SIIOQORRPSLLSEFOP--GNERSOELHLR-----PESHLYPEL 95			
Db	206	HQSGDAPSVFGATNIQQSPSPASTKQSSPYEDKDKKEKSAVRPSPSPERSSTGP 265			
Qy	96	GKSEWEIFESKRPLELLDPDLL-----RPSLLATQOPAGSDLL--TKDRSLT-GKLE 146			
Db	266	P-----APTELLVEQHDDSPRLAAI--PSSQFPVPSSSEASTRGCS 307			
Qy	147	PVSPSPSPHTDPELELVPRLSKEBELIQNMDRVREITVMVEQQISKLKKQKQOQLEEEAAK 206			
Db	308	PKSPEKPPQSTSS-ESCPP-----SPQPTKVSRRHASSPESILKP 345			
Qy	207	PPEP--EKPVPFPPIESKHSVLQIYIDENRKKABAAHRIIEGLGPQVELPLYNQPSDTR 264			
Db	346	TPAPGSRREISSSP-TSKNRSHGRAKDKSHST-PSHRAGRSRSPATKRGSRSRSTPTK 403			
Qy	265	OYHENIKINQAMRKKILIFYKRRNHARKQKQFCQRYDQLMEALEKKVRIENNRRA 324			
Db	404	RHRSRSRSPQ-----WRRSRAQRWKG-----SRSPQRRG 433			
Qy	325	KESKVREYERQFPEIRKQRELQERMOSRVQQRGSLGMSAARSEHEVSEIIDGLSEQEN 384			
Db	434	RSR-----SPQPCWSRSRNTQRRGRSARRGRSHSRSPATR-----GRSRRT 478			
Qy	385	LEKQMRQLAIVPMLYDADQOQRIKFINNGLMADPMKVYKDRQVNMNMWSEQEKTFREKF 444			
Db	479	PARRGRS-----RSTPARRRSRSTPARRRSR 506			
Qy	445	MQHPKNEGLIASFLERKTVASCVLYYLTKKNENKSLVRSYRRGKSQOQOQOQOQOQ 504			
Db	507	SRTPARRRGRS-----RTP-----TRRSRTRSPVRR--RSRGRSQARRSGRGRSR 551			
Qy	505	QOQOQOQPMPSRQBEKQBEKEAEKEEKEPEVENDKEDLLKEKTDGSDNDEKAV 564			
Db	552	TPARRSGRSRSTPARRGRSRSRTPARRSAR-----SRSRTP 588			
Qy	565	ASKGRKTANSQGRKGRITRMANEANESEAITPQOSABLASMELNFSRRWTEEMETAK 624			
Db	589	ARRGRSRSRTPARRRSR--SRSLVRGRSHSRTPQRRGRSGSSSRKKNKRTSQ-----640			
Qy	625	KGLLEHGRNWSAIARMVGSKTVSQCKNIFYNYKKRQNLDEILOQHLKWEKERNARR---681			
Db	641	-----RRSRNSSPPEMKKSHVSSRSRSLSS-----PRSKAKSLRRSL 679			
Qy	682	-----KKKKAPAAAEEAAPPVVEDEMEAGVSGNEEMVEAEALHASGNEVPRG 734			
Db	680	GSSPCPKQKQSTPTERSRSGSPPKQKSKTPPQSRSNSPQPKV-----SGTP-PR-730			
Qy	735	ECSGPATVNNSSDTESTPSP-----HTEAAKDTQNGPKPPATLGADG-----PP--779			
Db	731	-----PGSVTNMQADECTATPQQRSHSBSPP-GEVKSRTPRQSCSGSSPRVKSSTPPRQ 785			
Qy	780	-----PGP-----PTPRTSRAP-----IEPTPASEA-----802			
Db	786	SPSRSSSPQPKVTIVISPRGRSHSSSSSPSRVTSRTPQKRSRSVSPCPKVDRLRHSR 845			
Qy	803	-----TGAP-----TPPPAPSPSAPPVVPVPEKEEBETAAA-834			
Db	846	SRSSSPDSKMLGTPFLRHSGSTSPYPKSMLQTPPDQNLGSKSP--CPQKSRDPTGSSG 903			
Qy	835	-----PPV-----BEGEQKPPAAEELAVDTGKAEEPVKSECTEEAE 871			
Db	904	SFHLCPGVTPESSIVRGESCFASPVQKGHTQTPWDTSSPEVMQTQVESPLQSKSQSTP 963			
Qy	872	EGPAKGKDAEAAAEATGECALKAEEKGSGRATTAKSGAPQDS-----DSSA 919			
Db	964	KGSLSRSSSPVTELTARSPVKQDKSEISTD--PKLKGMSPEQSKTKPDSSIYPLVDSKS 1021			
Qy	920	TCSADEVDEAFGGDKNRLLS-----PRSLLTPTGDPANANASPOKPLDLKOLKORAA 971			
Db	1022	FLVQSRLEPSELKELGLIQEDVASSCIPRDK-PSPTQDRPESSTVLKVTVRVLKERSG 1080			
Qy	972	AIPPIQTVKHEPREDAATPKP-----APPAPPPQNLQOPESDAPQOPG--1016			
Db	1081	AGSP-----POGRDQKSLLPNSQODELMEVKEBQPLSQVLPSLSEH--KEMPGSN 1130			
Qy	1017	--SSPRGKRSRPPADKEAFAEAQKLPGBP--PCWTSCLPFPVPVPREVIKASPH-AP 1070			
Db	1131	IESSEPEVEER-PAVLSALDOSQOSPASKAAGTFAVASCW-----SGPQVSP 1174			
Qy	1071	DPSAFSAPPHPPLPLGLHDTARVLPRLPRTI-----SNPPPLISSAKHSPSLERQI 1122			
Db	1175	EKELSHSP-----PRENSFESSLEFPKNSGPVSEVNTGFSPEVKEBLN 1217			
Qy	1123	GAISQGMVQLHVPYSEHAK-----APVGP-VTMGLPLPMDPKKLAPFGVGKQEOQLSPR 1175			
Db	1218	GSFLNQTEADPSVDMKQSQSRSSRSELSEVVEVKVGLFSSQKVSSPVLETVQOORTPSR 1277			
Qy	1176	--CQAGPPESLGVPTAQEAASVLRGTALGSVPGGSIITKGIPSTRVPSDSAITYRGSITHG 1232			
Db	1278	ERSSASAPELKDGILPRTPS-----RRSRGSGSPGLRDGSGTFS-----RHSLSGS 1322			
Qy	1233	TPADVLYKGTITRIIGEDSPSRDLDRGE--DSLPGKHVYIEGKKGHVLSYEGGMSVTOCS 1290			
Db	1323	SPG-----MKDTPQTPSGRSECDSSPEPKALPOTPRAR-----SHSPSS 1362			
Qy	1291	KEDGRSSGPPHETAAKRYIDMME-----GRVGRAISSASIEGLMGRATIPPER----H 1340			
Db	1363	PERNNKSVTPQRRSGSGSESSVEQKNLARTSPQGRSRSGSQELDG-KPSASQPQRSESDS 1421			
Qy	1341	SPHH-----LKEOHIRGSIITQGI PRSVYEAQEDYLRRKAKLLKREGTPPPPPPSRDL 1393			
Db	1422	SPDKPKTRTPLRORSHSGS-----PEVDSKSRHSPLRSRSGSGSEPMKDK---1467			
Qy	1394	TEAYKTOALGFLKLPKPAHEGLVATVKEAGRSIHIPREELRHTTPELPLAPRLKEGSTIQ 1453			
Db	1468	-----PRVLQRAQSGTSSPE-----HKIP-----APRALPHRS-RS 1498			
Qy	1454	GTPLKYDTGASTTQSKKHVRSLLIGSPGRTFPVPHLDVMDADARALERACVSELSKSR-P 1512			
Db	1499	GSSSK-ERGPSPEGSSSESSSEPEHAPKSR-----ARRGRSSIEPTKSRTP 1545			
Qy	1513	GTASSSGSITARGAPVIVPEL-GKPROSPLTYEDHGAPFAGHLPRGSPVTWREPTPRQE 1571			
Db	1546	PRRSRSS-----PELTRKARVRRSRSSASSP-----EIRSRTPRRRS 1587			
Qy	1572	GSLSG-----SKASQDRKLTSTPREIAKSPHSTVPHEHPHI-----1608			
Db	1588	PSVSSPEPTKESRRRRRSVSSPRTKTSRRGRSPSPKPRGLQRSRSRREKTRTTRR 1647			
Qy	1609	-----SPYELLRGVSGVDLYRSHIPLAFDPTSP-----RG---I 1641			
Db	1648	RDRSGSQSTRRRQRGRSRSVTRRRRGSG--YHSRSPTRQESSRSTRRRRGRSRT 1704			
Qy	1642	PLDAAAAYLPRHLAPNTVPHLYPPVILRGYPDTAALENRQTTIINDVITSQOMHNAT 1701			
Db	1705	PLTS-----RKGRSRTSPAPWKRSRSRASP---ATHRRSRITPLISRRRSRSTSP 1754			
Qy	1702	AMAOADMLRGLSPRESSALNVAAGPRGIIDLQVPHLPVLVPPTEGTATAMDRLAYL 1761			
Db	1755	VSRRS--RSVNRRRSRSRASPVRRR--SRSRTP-----PVTRRRSRST 1796			
Qy	1762	PTAPQPFSSRHSSPLSPGGPHTLTKPTTTSSSRERDRDRDRDREREKSIITSTTV 1821			
Db	1797	PTRRR--SRSTPVTTRRRSRSTPPVT-----RRRSRSTSPVTRRRSRSTSPVTRR 1848			
Qy	1822	EHAPIWPGTEQSGSGSGSGGGGSSR--PASHSHAHQHSPISTPQD-----ALQO 1873			
Db	1849	RSRSTSPVTRRRSRSTSPVTRRRSRSTPPATRRRSRSTPLLRKRGRSRSPALAIR 1908			
Qy	1874	RPSVLHNTGMKGIIITAVEPSKPTV-LRSTSTSPVSPAATPPATHCPLGGLDGV----1928			

Db	10765	KVPEPKVPEEKIRISITKRE---KQVTEPAKVFPMKE-----RVVAEEKVPVRKE	10816
Qy	898	GGSGRAITAKSSGAPOQSDSSATCSADE---VDEAGGDKNLLSPRPSLLT-----	946
Db	10817	---VAPVRVPEPKLEPEEVAPEEVEVTHVEEYLYEIEEYIIEEDEFITEEVEVVPV	10872
Qy	947	-PTGDPRANASP---OKPLDLKQKORAAAIPIQVTKVHEPPREDAAATKPAAPPPP	1001
Db	10873	IPVKVPEVPRKVPPEEKVPVPKKE---APPAKVPEPKPEEKVPVLIPEKSKPPP	10928
Qy	1002	PONIOPESDAPQOGSGSPRGKRSAP-PAKFAFAEAAQKLPQDPPCWTSGLPFPVPPR	1060
Db	10929	---AKVPEV- KRPVPEKVPVPEKVEAPPKAPKPEVPKK-----VPEK	10970
Qy	1061	EVIKASHAPDPSAFSAPPGH--PLPLGLHDTARPVLPRPPTISNPPPLISSAKHPSVL	1118
Db	10971	KVPVPAPKVE-----APPAKVPEVKLIPEEKKTVPVKVEAPPKVPKKEP---	11021
Qy	1119	ERQIGALSQMSVOLHVPYSEHA--KAPVGPVTWGLP-----LPMDPKKLAPSGV--KQ	11169
Db	11022	-----VVPVVALQEEEVLPEEIVPEEVLPEEERVLPEEVEVLPEEVEVLPEE	11071
Qy	1170	BOLSPRGOAGPPESLGVTAQASVLRGTALSGVPGGSITKGI-PSTRVPS-----	1223
Db	11072	BEIPPEEVEVPPEEYVPEE---FVPEEVLPEVKVPVPAVPEIKKV	11121
Qy	1224	TVRSGI---THGTADVLXGTITRIIGEDSPRLDRGREDSLPKGHVIEYEGKKGHLS	1279
Db	11122	TEKKVVPKKEAPKVV-----PEVPKVEKR-----LILPEEEVLP	11161
Qy	1280	YEGGMSVTQCKSGRSGGPHETAAPKRTYDMMGVRGVAISASIEGLMGRAIIPER	1339
Db	11162	VE---VTEPEEPISEEEIPEEPPSIEEVEAVPRVPEVIKAVEA--PTFVPKVV	11215
Qy	1340	HSPHLKQEHRTGSIQTGIPRSYVDAQEDYLREAKLLKREGTPPPPPSRDLTEAYKT	1399
Db	11216	EAP-----PAKVKKIPKEKVPVPVQ--KKEAPPAKVPEVKVPKVKVLP--KK	11262
Qy	1400	QALGPLKPKAHEGLVA-----TVKE-----AGRSIHEIPREELRHTPELPLAPRL	1446
Db	11263	EAVPAKGTVLKEKVSVAFRQEVVVVKERLEWVAEVEIEPEEPEFHEV-----EYF	11317
Qy	1447	KGSGITQGPLKVDTCASITGSKKIDVRSLSIGSGRTFPPVHPLD-----VMADAPALERA	1502
Db	11318	ERGEF-----HEVEEFIKLEQHRVEEHEHRVKVRVIEVFEEAVE	11358
Qy	1503	CYEESLKSRTGATASSGGSIARGAP--VIVPELGPQROSP-----LTVEDHGAPFAGHL	1554
Db	11359	VFEKP-KAPPKGEPISEKLIIPPKPTKVPVRKEPPAKVPEPKIVVEEKVRVP--EE	11414
Qy	1555	PRGSPVTMBREPTRLQESLSSKASQDKLSTPREIAKSPHSTVPEHHPIPSYEH	1614
Db	11415	PRVPTKVPDLVLPKEVPEKVPVPAPKKEAPPKVPAPKEVVPKVV-PVPPPK--	11471
Qy	1615	LRCVSGVDLYRSHIPLAFDPTSTPRGIPLDAAAAYLPRHLAPNTYPHLYPYLIRGVP	1674
Db	11472	-----KPEVP-----PTKVPE-VPKAAVPEKKVPRAPIPKPESP--PPEVPEA-P	11512
Qy	1675	DTAALENROTIINDYITISQMHNTATATAQARADMLRGLSPRESSIA-LNYAAGRGIID	1733
Db	11513	KEVPEKK-----VPAAPPKKPEVTPVKVPEAPKEVVP	11545
Qy	1734	LSQVPHLVLPVPTPGTATAMORLAVLPTAPOPFSRSHSSPSLPGGTHLTKPTTSS	1793
Db	11546	EKKVP-----VPP-PKKPEVPTKVPKEVPAVPEKKVPEAPIPKPEP-----PPVFEE	11595
Qy	1794	SERERDRDRDRREKESILSTTTVEHAPTRWPGTEOSGSGSGSGSGSGSSSRPAS	1853
Db	11596	PEEVAEEPPAEVVEEPEAPQV-----VPPKVPVEKKA-----PAV	11636
Qy	1854	HSHAHQSPISPRTDALQORPSVLHNTGMKGIIITAVEPSK-----TVLRST	1901

RESULT 39
O63460
ID O63460 PRELIMINARY; PRT; 5027 AA.
AC O63460;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Proline-rich protein (Fragment).
GN PRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Ventral prostate;
RX MEDLINE=89066721; PubMed=3198617;
RA Hemschoote K., Peeters B., Dirckx L., Claessens F., De Clercq N.,
RA Heyns W., Winderickx J., Bannwarth W., Rombauts W.;
RT "A single 12.5-kilobase androgen-regulated mRNA encoding multiple
RT proline-rich polypeptides in the ventral prostate of the rat.";
RL J. Biol. Chem. 263:19159-19165(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Ventral prostate;
RX MEDLINE=92250652; PubMed=1577819;
RA De Clercq N., Hemschoote K., Devos A., Peeters B., Heyns W.,
RA Rombauts W.;
RT "The 4.4-kilodalton proline-rich polypeptides of the rat ventral
RT prostat are the proteolytic products of a 637-kilodalton protein

QY 1610 PYEHLRGVSG-----VDLY-----RSHIPLAFTSIPRG-----IPLDA 1645
Db 2400 IYLHAEPPPTGPFVEPPDLFLTKTKSKPVQSGPRQVDKSPKEMFTQSPYEESVLPQA 2459
QY 1646 AA--AYVLPRLADNP-----TYPH--LYPP-----YLIRGYPDTAALENRQTII 1686
Db 2460 EQBESRAPHMSLQPLDQDLTLSSHPHGWIHTHPNTPDKIYLHYAEPPTGPFVEPPDLF 2519
QY 1687 NDYITSOQHNTATAMAQADMLRGLSPRESSALNAYAGPRGIDLSQVPHLPVLVPP 1746
Db 2520 FLRTTKSKPAQWTPTQMAKSPPEMVLSPENKETV--FPAQKGQESISPPHMSLQPLD 2577
QY 1747 TPGTPATAMDLAVLPAPQPFSS--RHSSPLSPGCP-----THLKPTTSS 1793
Db 2578 QDLTFSSHPH--GWIPHPNTHGKIYLHYAEP--PTGPFVEPPDLFLTKTKSKPVQSP 2633
QY 1794 SERERDRDRDREREKSILTTTVEHAPIWRPGTEQSSGSSGSSGGGSSSRPAS 1853
Db 2634 RQIDKSPKEVFTQSEESVL-----PQAEGQEE----- 2664
QY 1854 HSHAHQHSPIPRQD--ALQORPS--VLHNTGMKGII-----TAVEPSKPTVLRS 1900
Db 2665 -SRAPPMSLQPLDQDLTLSSHPHGWIHTHPNTHGKIYLHYAEPPTGPFVEPPDLFLTK 2723
QY 1901 TS-----TSSP-----VRPA-----ATPPP-----A 1916
Db 2724 TKSPVHSGPRQIDKSHKEMFTQSEESVLPQAQESRAPHMSLQPLDQDLTL 2783
QY 1917 THCPILG-----GTLDGVYPTLMEPVLPLP-----KBAPVARPERPRADTGH 1957
Db 2784 SH-PHGWIHPNTPDKIYLHYAEPPTGPFVEPPDLFLTKTKSKPVQSGPRQIDKSPKE 2842
QY 1958 AFLAKPPARSGLPASPSPK--GSEPRPLVPVSGHATARTAKNLAPHA--SPD--- 2010
Db 2843 MFTQSEYESVLPQAQDQESRAPHMSLQPLDQDLTLSSH--HGMIPHPNTPDKIY 2901
QY 2011 -----PPAPPASADP--HREKTSKPFISQLELRLSLGVHGS-----SVSP 2050
Db 2902 LHYAEPPTGPFVEPPDLFLTKTKSKP-----VHGPRQIDKSPKEVFTQSP 2948
QY 2051 EGVEPVSVPSPSLTHDKLPKH--LEELDK-----SHLEGEL--RPKQPGVKLGGEAA 2101
Db 2949 EYESVLPQAQDQESRA--PPHMSLQPLDQDLTLSSHPHGWIHTHPNTPDKIYL----- 3002
QY 2102 HLPHLRPLPEQSPSSPLLOTAPGVKGQRVVTLAQHISEVITQDYTHHPQQLSAPLPA 2161
Db 3003 HYAEPPTGPFVEPPDLFLTKTKSKPVHSGPRQIDKSPKEVFTQS-----BEYESVLPA 3057
QY 2162 PLYSPFGASCP-----VLDLRPPSDLYLPPPD 2189
Db 3058 QAEDQESRAPHMSLQPLDQDLTLSSHPHGWIHTHPNTHGKIYLHYAEPPTGPFVEPPD 3117
QY 2190 -----HGAPARGSP-----HSEGGKRSPEPNKTSV--LGGGED-----GIEPV- 2225
Db 3118 LFFLTTRTKSKPVQWPTQIDKREEMVQSPENEEADIPRHGDGQDELNPPHIIISLQPLD 3177
QY 2226 -----SPEG--MTEPGRSRAVYLLYRD--GEQTEPSRMGSKSPGNTQPPAFFSKL 2275
Db 3178 QBLTLTLSSQPHGVVPPPTPDKIY--LHYAEPPTGPFVEP-----PDLFFLRT 3223
QY 2276 TESN-----SAMVSKKQEIKNKLNTHNRNEPEYN--ISQPGTEIFNMPAITGTGLMT 2326
Db 3224 TKSKPVQGTPTQAKSPKEMVFO-----TPYKEAVLSAPGQDET- 3266
QY 2327 YRSQAVQEHASTNMGLEAIRKALMGKYDQ--WEESPP-----LSANAFN----- 2370
Db 3267 -----SSPNMSLQSDQBLTMSQPHGLPHPPNTHGKIYLHYADPPTGPFVEPPD 3317
QY 2371 ---LNASASLPAAMPTTAADGRSDHDLTSPGGGKAKVSGRPSRKAKSPAPGLASGRDP 2427
Db 3318 LFFLTTRTKSKPVQGTQIMD-----KSPKQGL----- 3345
QY 2428 PSVSSVHSGDCNRRRTPLTNRVME--DRPSSAGSTPF-----PYNPLINRLQAGVNASPP 2480

Db 3346 ----PAHQEAQDESSPPNMSLQPDQELSLSSQPHGWIHTHPNTPGKIYLH---YABPP 3398
QY 2481 -----PPQL-----PAGSPLAGPHAWDEEPKPLLCQYETLSDSE 2517
Db 3399 TGFVVEPPDLFLTKTKSKPLAGSPQMAKSPKEMPS---QTLHKE 3442

RESULT 40
O88737 PRELIMINARY; PRT; 3942 AA.
AC O88737;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Bassoon.
OS Mus musculus (Mouse).
GN BSN OR BASSOON.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129 SVJ;
RX MEDLINE=98345363; PubMed=9679147;
RA Dieck S., Sanmarti-Villa L., Langnaese K., Richter K., Kinder S.,
RA Soyke A., Wex H., Smalla K.H., Kampf U., Franzer J.T., Stumm M.,
RA Garner C.C., Gundelfinger E.D.;
RT "Bassoon, a novel zinc-finger CAG/glutamine-repeat protein selectively
RT localized at the active zone of presynaptic nerve terminals";
RL J. Cell Biol. 142:499-509(1998).
DR EMBL: Y17034; CAA76598.1;
DR EMBL: Y17035; CAA76598.1; JOINED.
DR EMBL: Y17036; CAA76598.1; JOINED.
DR EMBL: Y17037; CAA76598.1; JOINED.
DR EMBL: Y17038; CAA76598.1; JOINED.
DR PIR: T42730; T42730.
DR MGD; MGI:1277955; Bsn.
DR GO; GO:0045202; C:synaptic junction; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR InterPro; IPR008899; Znf_piccolo.
DR Pfam; PF05715; Zfpiccolo; 2.
SQ SEQUENCE 3942 AA; 418739 MW; 150267E636C4DACB CRC64;

Query Match 3.8%; Score 504; DB 11; Length 3942;
Best Local Similarity 19.1%; Pred. No. 2.1e-15;
Matches 590; Conservative 324; Mismatches 1122; Indels 1052; Gaps 136;

QY 55 PGSIQORRRPSLLSEFPQGNERSQELHURPESHVYLPGLKSEMEFIESKPRLELLP 114
Db 557 PQGLGQPSGLPAKAS-----POATKASPOATKASPOATKASPOATKAS 601
QY 115 DPLLRPSPLATGQAPGSEDLTKDRSLT-GKLRPVSPSPPHDPELELVPRLSKEELI 173
Db 602 ----QAKPLRATPEKTSSTSSAQEKATPAKAPV-PKPPET-----TVPGTGPKAK-- 649
QY 174 QNMDVRDREITMVEQQISLKKKQQQLLEBAAPKPEPEKPVSPSPPIESKHSRLVQIIYDE 233
Db 650 SGVKRTDPAIPVVKP-----VPEAPKGEAEPEVPKPYSDLSRS----- 689
QY 234 NRKKAEAHRILEGGLGQVVELPLYNQPSDTFQHENIKINQAMRKLLILYFKRNHARKQ 293
Db 690 -----PQ-----SLSDTGYSDGVSQS----- 708
QY 294 WKQFCQRYDQMEALEKVKVERIEN--NPRRRAKESKREYVEYKQFPEIRKQRELOERM 350
Db 709 -----EITGVQVEQVQLDSAGVTGPRPSPSELHKVGSLSRPSLEAQAVAPSAE 758
QY 351 QSRVQSGGSLMSAAKSEHEVSEIIDLSEQENLEKQMRQLAVIPMLVDADQORIKFI 410
Db 759 WSKPSPSSSSAVEDQKRRPHSL-----IMPEAFDSDELDGIL 797
QY 411 NMGLMADPMKVKYKDRQVMNMNWSQEKETPREKFMQHPKPNGLIASFLERKTVAECLVY 470

Qy 798 -----PASEATGATP-PPAPSPSAPPVVPVPEKEEETAAA-----P 835
Db 6441 YERVEEHEVITEPEKPIPVKVPVEEPVTPKAPPAKVLKAVPEEKVPVPIPKLKP 6500
Qy 836 PVEEGEQQKPAARELAVDTKAEPEVKSECTEAEAGPAKGADEAAEATGALKAOK 895
Db 6501 PKVPPEPKVFEKIHISITKRE---KEQVTEPAKVPMPKP-----RVVAEKKVPVR 6552
Qy 896 KEGSGRATTAKSGAQODSSATCSADE-----VDEAEGGDKNLLSPRSLIT----- 946
Db 6553 KE-----VAPPVRVPEVKELPEEVAPEEVEVTVHVEEVLVEEVEEYTHEEPEFITEEVV 6608
Qy 947 ---PTGDPRANASP-----QKPLDLKQKORAAAIPIQVTKVHPREDAAPTKAPAP 999
Db 6609 PVIKVPVPEVRKVPPEEKVPVPEKKE-----APPAKVPVPEKVPPEKVPVLPKKEKP 6664
Qy 1000 PPQNLQPEDAPQOPGSSPRKGRSPAP-PADKEAFAAEAKQLPGDPWCWTSGLPVP 1058
Db 6665 PP-----AKVPEVP-KKVPPEKVPVPEKVPVPEKVPVPEKVPVPEKVPVPEKVP 6706
Qy 1059 PREVIKASHPADPSAFSAAPGCH--PLPLGLHDTARPVLPRPTISNPPPLISSAKHPS 1116
Db 6707 EKKVPVPAKPV-----APPAKVPVPEKVLIPPEEKVTPVPEKVPVPEKVPVPEKVP 6759
Qy 1117 VLEROIGAIQSGMSVOLHVPYSEHA--KAPVGPVTMGLP-----LPMDPKKLAPFGV-- 1167
Db 6760 -----VPVVALQPEEVLPEEVEIVPEEVLPEEVEVLPEEVEVLPEEVEVL 6807
Qy 1168 KQQLSPRQAGPESLGVTQAQASVLRTALGSVPGGSITKGI-PSTRVPS-----DS 1221
Db 6808 EEEIPEEVEEVPPEEYVPEE-----FVPEEVLPEKVPVPEKVPVPEKVP 6857
Qy 1222 AITVRSI---THGTADVLTKYITRIIGEDSPSLDRGREDSLPKGHVYEGKGVH 1277
Db 6858 KVTBKKVVPKKEAPKAV-----PEVKKVEER-----IILPKEEV 6897
Qy 1278 LSYEGMSVTCQSKEDGRSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAPP 1337
Db 6898 LPVE---VTEPEEPISEBIEPEEPSIEVEEVAAPPVPEVVKAVPEA--PTVPK 6951
Qy 1338 ERHSPHLKHQHRIGSITQGIKPSYVEAQEDYLREAKLLKREGTTPPPPPSADLTYAY 1397
Db 6952 KVEAP-----PAKSKKIPKEKVPVQV--KEGAPKVPVPEKVPVPEKVPV 6998
Qy 1398 KTOALGLKLKPAHEGLVA-----TVKE-----AGRSIHETPRELHRTPELPLAPR 1444
Db 6999 KKEAVPAKGRTVLEEKVSVAFROEVVVKERLEEVVEAEVEEIPPEEPEHEV--E 7053
Qy 1445 PLKEGSITOGTPLYKDYTGASTGSKGHDVRSLSIGSPGRTPPPVHPLD-----VMADARALE 1500
Db 7054 YFEBGEP-----HEVEEFIKLEQHRVEEHRVKEVHRVIEVFEAE 7094
Qy 1501 RACVEESLKRPGTASSGSGSIARGAP--VTVPLGKPRSP-----LTVEDHGAEPAG 1552
Db 7095 VEVFEKP-KAPPKGEISEKIIIPKPKPTKVPVPEKVPVPEKVPVPEKVPV 7150
Qy 1553 HLPKGSVTVREPTPLQEGSLSSKASQDKLTSTPREIAKSPHSTVPEHHPHIPISYE 1612
Db 7151 BEPVPTKVPVLPPEVPEKVPVPPKAPPAKPPKVPPEKVPVPEKVPVPEKVPV 7209
Qy 1613 HLLRGSGVDLYRSHIPLAFDPTIPRGIPLDAAAAYVLPRLHAPNPTYPHLYPYLING 1672
Db 7210 -----KEVP-----PTKVPB-VPKAAVPEKVPVPEKVPVPEKVPV 7248
Qy 1673 YPDTAALENRQTIINDYITSOQMHNTATAMAQADMLRGLSPRESSLA-----LN 1723
Db 7249 -----LPVLVP-----EPESPSAPPKKEVPEVVR 7267
Qy 1724 YAAGPRGIIDLQVPH-----LPVLVP-----PTGCTATAMDLAYL 1761
Db 7268 VPEVPKEVPEKVPVPAAPKPKPEVTPVKVPEAPKEVVPVPEKVPVPEKVPV 7327

Qy 1762 PTAPQPFSSRHSSPLSGGTHLTKPTTTSSRERDRDRDRDREREKSIIITTTV 1821
Db 7328 PKVAVPEKVPKPEAIKPKPESP-----PPEVPEPEVEALEEPPAEVVEPEPAAPQVTPV 7383
Qy 1822 EHAFIWPPTGTEOSSGSSGSSGSSRSPASHAHQHSPISPRTQDALQORSVLHNT 1881
Db 7384 PKNEV-----PEKKAPAVAKKPELPEV-----KVPEVPKEV 7415
Qy 1882 GMKGIIITAVPSKP-----TVLRSTSTSSVVRPAATPPATHCPLGGTLGDGVY 1929
Db 7416 VPEKKVPLVVPKKEAPPAKVPVPEKVPVPEKVPVPEKVPVPEKVPVPEKVPV 7463
Qy 1930 PTLMEPVLLPKBAPRVARPERPRADTGHAFIAKPPARSGLEPASPSK-----GSEPRP 1983
Db 7464 EVPKKPVL--EEKDAVPVPERAESPPPEV--BEPEIAPBEETAPBEKVPVVAESEP 7519
Qy 1984 LVPPVSGHATITARTPAKNL-----APHASDP-----PAPPASADP 2021
Db 7520 EVPP-----PAVPEBKPIIPEKKVPVVKPEAPPPEPEKVPVPEKVPVPEKVPV 7575
Qy 2022 HREKTQSKPFSI-----QELRLSLGYHGS-----SYSPGVEPVS 2057
Db 7576 PKEDVKEKIQLKAIKKKVPENPOVPEKVELTPLKVPGGKKVKRKLPERKPEKEVV 7635
Qy 2058 PVS-----SPSLTHDKGLPKHLEELDKSHLEGELPKQGPVK-----LGGEAAHLPLR 2107
Db 7636 LKSVLRKRPEEPEKVPKLEKVKKPAVP--EPPPKPVEEVEVPTVKRERKIPEPT 7692
Qy 2108 PLPESQSSSPLLOTAGVKGHQRVTVLAHQHISEVITQDTRHHHPQOLSAPLPAPLSFP 2167
Db 7693 KVPEIKP-AIPLPAPEKPKPEAEVKT-----IKPPVVEPEPTPIAAVTVVVGKK 7743
Qy 2168 G-ASCPVLDLRRPSPDLVPPDHGAPARGSPHSEGGKSPSPNKT--VLGGGEDGTEP 2224
Db 7744 AEAKAPKEAAKPK-----GPIKGV-----KTPSPIDEAERKLRPGSGGKP 7787
Qy 2225 -----VSPPEGMT----- 2232
Db 7788 PDEAPFTYQLKAVPLKFVKKEIKDILTESBFGSSAIFECVLSPTAITTWKDGSNIRE 7847
Qy 2233 EPGHSRA-----VYPLLRYDGEQTEPSRSGSKSPGNTS-----QPPAFSKLTES 2278
Db 7848 SPKHFRTADGDKRKLHIIDVQLSDAGEYTCVLRGLNKEKSTAKLVVEELVRFVKTLEE 7907
Qy 2279 NSAMVSKS---KQINKLNLTHNRNPEYNIOPGTEIFNMPAIG 2321
Db 7908 EVTVVKQPLVLSCELNKRDDVVRKDGKIVVEKGRIV---PGVIG 7951
RESULT 42
Q9N3Y8
ID Q9N3Y8 PRELIMINARY; PRT; 2344 AA.
AC Q9N3Y8, 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Y40C5A.3 protein.
GN Y40C5A.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RN Science 282:2012-2018 (1998).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;

RA Kalicki J., Randall J.;
 RT "The sequence of C. elegans cosmid Y40C5A.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC024772; AAF60538.1;
 DR WormPep; Y40C5A.3; CE24257.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0015992; P:proton transport; IEA.
 DR InterPro; IPR000194; ATPase_a/bcentre.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
 SQ SEQUENCE 2344 AA; 245902 MW; 7D8696D48D0ADDDB CRC64;

Query Match 3.8%; Score 503; DB 5; Length 2344;
 Best Local Similarity 19.2%; Pred. No. 1.2e-15;
 Matches 461; Conservative 275; Mismatches 865; Indels 796; Gaps 109;

Qy 515 SSOEKEDEKEKEBAEKEEKEPVENDKDLLEKTDGSDENDEKEAVASKRKTANS 574
 Db 218 STTEETTEETTTVSTNPPTPELSDD-----PTDNDATDIDDK----- 257

Qy 575 QRRKGRITRSMANEANSEEAITPQQAELASLMELNESSRWTEEMETAKGLLEHGRNW 634
 Db 258 -----IVKVESSTYAEPTVAPFFENTVTSVFEPDDADFDKPVKSKSTLPNGINS 310

Qy 635 SAIRMVGSKTVSQCKNFYFYKQRQNLDEILOQHLKMEKERNARKKKKAPAAASEEA 694
 Db 311 KRSTSYANPLTVDNA-----AABSRNKLSCLPAPGTV 344

Qy 695 APPVVEDEMEASGVSENEEMVEEARALH--ASGNEVR-----GECGPATVN-NS 745
 Db 345 AVPP-----KEENLPIEPLPSVAGYQPPRTLPVAPGDPSESTFVQIHE 389

Qy 746 SDTESIPSPHTEAAKDTQNGKPKPATLIGADGPPGPPPTPRRTSRAPIEPTPASEATGA 805
 Db 390 TETSPASPSPSSSVTVIPRVALPASVSPSPASPNVPTIILSAPAK-----SPATYG 445

Qy 806 FTP-----PPAPPSAPP-----PVVPKE-EKEEETAAPPVEREEQKPPAA----- 848
 Db 446 PAPEPSRPAPVPAPAPSVIPTTSQSPQKQTPAQTPSPAKPIPEGPAQSPNAVKTYS 505

Qy 849 -----RELAVDTGAE--EPVKSECTEAE-----GPAKGD 879
 Db 506 GSSPAAPRVPEMIEVAPGNVEKTPDQTNQVPETSHENAPAPQPSPPSVYGTPTSEP 565

Qy 880 AEAATATAGALKAR-----KKEGSGRATTAKSS-----GAPQDSD-- 916
 Db 566 AKPSMPGESPAQPPSPVAVNPSPQSPSSSGPIQTAPSPSPNAIPEGPPGPDVP 625

Qy 917 -----SSATCSADEVDEAGGDKNRLLSRPSL-----LTP 947
 Db 626 VMIALPPPKSPYNSGAGEDKGEVDQEVNPSER-----RPEISSISENSIENTSKTKTP 681

Qy 948 TGDP-----ANASQKPLDLKQKORA-----AAIPPIQVTKHE----- 983
 Db 682 TTVPKSTITGRMPSVNPSPDNLSNMEDGNAINQPSFNSLSETSITPASFGHDGKKQ 741

Qy 984 -----PPREDAA-----PTKPAPPAPPPQNLQSPESDAPOQSGSPGKSRSP 1026
 Db 742 EKLPGSKGPIPPRQESITGSTRHTRGSSPSIIPYATAPTSPPDKKPTIP---PTFP 798

Qy 1027 APPAKAEFAAQAQ-----KLQDPPCWTSGLPFPVPPREVLIKAS-----PHADPP 1072
 Db 799 TLSDVSDGIGQELRITRSPNPVALPSPRSKTLTLPSELPIINLPSARQSTSPASVENV 858

Qy 1073 SAFSVAPCPHPLPLGLHDARVLPVLPPTISNPPPLISSAKHPVLERQIGAIQSGMSVQ 1132
 Db 859 SNNSPSPSQSIPI-----ESTVIFGTPTSPISPMETRSTAN-----SSK 899

Qy 1133 LHPVYSEHAKAPVGPVTMGLPLPMDPKLAPPSGVKQ-----EQLSPRQAGPPESL----- 1184

Db 900 NHPVSSKISNRLPVSTGSKIHLTK--SPOTGAMENSGEVISEEPIGKPVSPGGTG 957
 Qy 1185 -----GVPTAQEASVLRGTALGVPGGSITKGIPISTR-----VPSDAIYTRGSIT-- 1230
 Db 958 AALPYTSPFKTQKSTFASTDVHSASAPSLPKLSHKNFEKATPPSISAAITTNVVD 1017

Qy 1231 ---HGTPA-----DVLKGTITRIICEDSPSRLDRGREDSLPKGHVYEGKXGHLVS 1279
 Db 1018 AFENSPAAPAVSTWEEVTPSGT-TASISQSQT-----PQPGGAVPP 1059

Qy 1280 YEGGMSVTQCKSKEDGRSSGP-PHETAAPKRYTDMMEGRVGRALISASIEGLMGRALPE 1338
 Db 1060 YALFPRTHQSSNSLSGTSSLPTSSSSAPVK-----VSSKTIEM--KTPPS 1105

Qy 1339 ---RHSPHLKEQHHRGSIITQIGIPRSVVEAOEDY---LRREAKLLKREGTTPPPPSRD 1392
 Db 1106 SVSSSGSLQIQEPTQAS-----SNVVEISESSPIVSSVSTWEEITSQSPASPTSQ 1159

Qy 1393 LTEAYKTQALGPKLKP-----AHEGLVATVKEAGRSIH---EIPREELRHTPELPLA 1442
 Db 1160 ILTAKSRPTGQGI VPPVYALPFRPHETL-STASPPASSLSASAKLPKSESQFMNAIP-- 1216

Qy 1443 PRPL-----KEGSIQGTPLKYDTGAST---TGSKKHQDVRSGLISGPGTFFPVHPLDVM 1493
 Db 1217 PSEFVDSGSGSGSGFGLQIPTTTDSSTPMDPSMEDYDVLTVPTAVTSGPVL-- 1272

Qy 1494 ADARALERACYBESLKRPGCTASSGSGSIARGAPVIVPE-----LGKPROS 1539
 Db 1273 -----TSNSRQSLITS-----LISSGKPATIPDMIMAWYTSQSSSKFSKSTAF 1317

Qy 1540 P-LTYEDHGA--FAGHLPRGSPVTRRPTPLQEGSIS-SKASQDRKLTSTPREIAK 1594
 Db 1318 PTLITQTEIKPTTVSSPAPKVPETKINQPTISSNIGLSKTEKESTIPREVEVP---SA 1374

Qy 1595 SPHSTVPEHH-----PHPSVPEHLRGGVGVLYRSHPLAFADPTSIP 1638
 Db 1375 SPBSTVTEHIDAETNSATIPTTGTTPPLKPKTKL--GLT-----SH-PSAIPPAIS 1425

Qy 1639 RG--IPLDAAAAYLPHRLAPNPT-----YPHLYPP----- 1667
 Db 1426 SKTLAPPVAPPTVTSNIAPISTHQSQOTRPTTTHRGITPPLAKPTIYPSLSQTGS 1485

Qy 1668 -----YLIRG-----YPTDALEN-----RQTI----- 1685
 Db 1486 SSPTPGTSSIIIVAGSRASSNYPTTASITDGSSEEBEENTRIPEETLSPASARLWT 1545

Qy 1686 ---INDYITQQMHNTATAMQADMLRGLSPRESSLA---LNVAAGPRGI----- 1731
 Db 1546 PDMISKYISSKAISTESVSS-AEQPSYLSNLPQSSSVASVKPPSSSVQNIETVPS 1604

Qy 1732 -----IDLQVPHLPVLVPTTGTPTATMDRLAYLPTAPQPFSSSHSSDPLSPGQPTH 1784
 Db 1605 YFPVEQSKSVTQAPIPASSPSAPSIPISTLPLVAPS-----SSSVSPSPSPASE 1655

Qy 1785 LTKPTTTSSERERDRDRDREREKSIILTSTTVEHAPIWEPGTEQSGSGSGSGG 1844
 Db 1656 -----STSTVSETP----- 1664

Qy 1845 GGSRRPASHASHAHOHSPISPT--QDALQORPSVLHNTGMKGIIITAVEPSKPTVLKSTS 1902
 Db 1665 APVNVNPSAHEAECEGYEPLSPQPAENHVEEYQNT---VDTIIEKS-----ATP 1714

Qy 1903 TSSVPRFAATPP---PATH---CPLGGLGVYPTLMPEVLLPKAEARVARPERPRADTG 1956
 Db 1715 APAPVTQAVQAPGPGVHEHRYEIPAPGPAPG--PAL-EPAPAPTSAPQIVEP----- 1763

Qy 1957 HAFIAKPARSGLEPASPSKSGSEPRPLVPVPSGHATIAITPAKNLAPH-HASDPDPAPP 2015
 Db 1764 ---LPPVQD--LPQOPTPEPEPLPIA-----TPAQOPTHTYGAQGNIVP 1805

Qy 2016 ASADPHREKTSQKPFSTQELSLGSLGVHSGSYGEGVEPVSPVSSPSLTHDKGLPK--- 2072

```

Db 1806 VTTAPVYVQETQAPI-----APSNPEVPQ--NPSQT--VSIPEPTG 1844
Qy 2073 HLEELDKSHLGEELRKPQGPVKLG-----GEAHLPLRLPESQPSSSPLLOTAP 2124
Db 1845 NIEEAE--HVDA--KPSYFGSSNNLEEDHEEGHIAELPALEPV--SVPTAARSL--P 1895
Qy 2125 GVKGHQRVVTQAHISEVITQDY--TRHHPOQLSAPLAPLAPYFPQASCPVLDLRRPPSD 2183
Db 1896 GESNYSN-----LEEDHDDKHLVGEPSVPFTQTSSGSSYSILEEDNQETN- 1943
Qy 2184 YLPPDHCAPARGSPHSGGKRSPEPNKTSVLGGEDG--LEPVSP-PEGWTECHSRSA 2240
Db 1944 -----NSGGYSYSPDKQVQVVPVTKPA--GGYRDGGMVENTAPQTTVQRP----- 1990
Qy 2241 VYPLLYRDGEOTSPRMSKSPGNTSQPPAFSKLTESNMAVSKSQEINKLKLTHNRN 2300
Db 1991 -HPKIVQVRPATKPYR--PRSPAPRIQRPY-----NPQPVVQRPQRPQRPQ 2041
Qy 2301 EPEYNISQTEIFNMPAITGLMTYRSQAVQEHASTNMGLEAIIRKALMGKYDQWES 2360
Db 2042 QPQ---POQPQ-----POQPQY-----IQRPALTLPFOQPQYP 2073
Qy 2361 PPLSN-----AFNPLNASASLPA-----AMPITAADGSDHTLTSPGG 2400
Db 2074 PPIFPQRPAPFIPFPQVQRPSPAGCGGOLFQGGGLCMPIFNPCQOQQQNNCGG 2133
Qy 2401 GKAKVSRGPRSKAKSPAPGLASGDRPPSVSVHSGECNRTPLTNVWEDRPSSAGST 2460
Db 2134 CGGCGSGGNSCGGCGNSGCGGCGGCGSSC-----CQPTSATSHAANKLNRRVAV 2188
Qy 2461 PFPYNPLIMRLOAGVMASPPPPGLPAGSGPLAGPHAWDERPKPLCSQYETLSDSE 2517
Db 2189 PL-----NNPTAGINDH-----LATQWESVSEGE 2212

RESULT 43
Q81Q71 PRELIMINARY; PRT: 2465 AA.
AC Q81Q71;
DT 01-MAR-2003 (TreeBrel. 23, Created)
DT 01-MAR-2003 (TreeBrel. 23, Last sequence update)
DT 01-OCT-2003 (TreeBrel. 25, Last annotation update)
DE CG323394-PA.
GN CG323394 OR CG10115.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC - Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Buesam M.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```

```

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Ye J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Buesam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradycky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003561; AANI2081.1; -.
DR FlyBase; FBGN0052394; CG323394.
DR GO; GO:0003824; P:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00455; AMP BINDING; 1.
DR PROSITE; PS50014; BROMODOMAIN 2; 1.
SQ SEQUENCE 2465 AA; 267809 MW; 716D22341B181188 CRC64;

```

Query Match 3.8%; Score 502; DB 5; Length 2465;

Best Local Similarity 18.8%; Pred. No. 1.5e-15;

Matches 530; Conservative 365; Mismatches 998; Indels 932; Gaps 130;

Qy 129 PAGESDITKORSLTGKLEPVSPPPPHPTDPELELVPPRLSKKEELIQN-----MDR 178

Db 28 PGSAQOQQQOQS-----QEPSSSSSSSGSTTTTASTVRSKKQAVLNSSGSSSTSKMADK 83

Matches		508; Conservative	245; Mismatches	926; Indels	923; Gaps	112;
QY	9	AQTWRAEPRYPH--SLSYVVOIA----	RTHDVGLEVOHHSRDYASHLSPGSI	1QPQ	62	
DB	212	AQRRSTCONSPRNVISGLPQANVLRKYS	ESG-LEVOHQ-----	PPR	256	
QY	63	RRRPS--LLSEFQGNERSQELHLPESH	SYLPGLGK--SEMEFIESKRRL--	ELLP	114	
DB	257	RAGPTVQLLLTLLPPLPAVGLVQLKLI	QLPWLGEVLPLOQDAGREMLSVNQVLP	316		
QY	115	DLLRPSLLATGAPGS-EDLTKD-----	RSLTG-----	143		
DB	317	STQRASSPETATKQSSPYEDKDKKES	ATRSPPSPERSTGPEPPATPLAERHGG	376		
QY	144	-----KLEVPSPSPHTDPELELVPP	RLSKKEELIQNMDRVREITWVEQOISKL	193		
DB	377	SQOPLATTPLSQEPVNPFS--EASPTR	DRSPK-SPEKLPOSS-----	SSESP	423	
QY	194	KKQOQLEBEAAKPEPEKPVSPPIES	KHRSVLQIYYDENKKAEEAHRILEGL	QPOVE	253	
DB	424	SPOPTKVSRHASSSPSPKPA--PAPG	SHREISSPTSKNRSHGRAKRDKSH	HTPSRR	480	
QY	254	LPLYNQPSDTRQYHENIKINQAMRKL	ILYPKRNHARK--QWKO-KFCORYDQ	LMEL	309	
DB	481	MGRSRSPATAKGRSRSTP-----	-----TKRSHSRSPQWRSSR	SAQWGR	524	
QY	310	EKKVERIENPRRAKESKVREYEQP	PEIRKQRELOERMOSRVGQSGSL	MSAAKSE	369	
DB	525	-----SRSPQGRGR-----	SPQPGWRSRNTQRRGRSARRGR	SHSRSPATR	570	
QY	370	HEVSEIIGLSOENLEKOMROLAVIP	PMLYADQORIKFINWGLMADPMKV	YKDRQVM	429	
DB	571	-----GRSRSTPARGRSRSTP-----	-----	589		
QY	430	NMSEQEKETFEKFMQHPKNGLIAS	FLERKTVAECVLYLYLTKNENYK	SLVRSYR	488	
DB	590	-----	-----ARRSRSTPTRRSRS	606		
QY	489	-----RRGKSQOQOQOQOQOQOQ	QOQOQPMPSQOEKDEKEKEKEKE	KEPEVENDKED	544	
DB	607	RTPARGRS--RSRTPARRSRTRSP	RVRRRSRSPARRSRS-----	647		
QY	545	LLKEKDDTSGEDNDEKEAVASGR	KRTANSQGRKGR1-TRSMANEANS	EEAITPOCSAE	603	
DB	648	--RSRTPARRG--RSRSTPARGR	SRSTPARGRSRSTPARGRSRSR	TPRRG--	700	
QY	604	LASMELNESSRWTEEMETAKGLLE	HGRNWSAIAIRWVG-SKTVSOCKN	FYFKRQNL	662	
DB	701	-----RSRSRSLVRRGRSHSR	TQRRGRSGSSSRKN-----	732		
QY	663	DEILQOQHLQWKEKERNARRKKK	APAAASEEAFPPVVEDEMEASGV	SGNEEMVEEAE	722	
DB	733	-----KSRTSQRSRN-----	-----SSPEMKSRISRRSR	SLSS--	764	
QY	723	ALHAGSNEVPRGEGSGPATVNN	SSDTESIPSPHTEAAKDTGONG	KPKPATLGADGPPGP	782	
DB	765	-----PKAKASRLSLRSLSGSS	-PCP-----KQKSQTPPRRSR	-GSSQPKAKS	808	
QY	783	PTPPRTRAPIEPTPASEATGAPT	PPAPPSAPPVPVPEKEEETAAP	VERGEE	842	
DB	809	RTPPRR-SRSSSSPPPKQK--SK	TPSRQSHSSSPHPKV-----	XSGTPRQSGI	855	
QY	843	QKPPAAEELAVDTGKAEPEPVK	SECTEAEAGPAKGDAAEAAATA	EAGLKAKEKGGSGR	902	
DB	856	TSPQANEQ-----SVTPQR	SRCSFESDPDELKS-----	-----RTPSRHSCGS	894	
QY	903	A-TTAKSGAQDSDSATSACSADE	V-----DFAEGDKNRLISPR	PSLLTPTGDPDR--	952	
DB	895	SPPRVKSSTTPRQSPSRSSSQ	PKVAILISPRQSHSGSS-----	SPSPSRVTSRTTPRS	950	
QY	953	ANASQPKPLDLKQKQ--RAAAI	PIPIQVTKVHEPPRED-----	AAAPTAP--	996	
DB	951	RSVSPCSNVESRLLPYRSHSG	SSPDTKVKPETPPRQSHSGS	ISPIYPKVKAQTTPGSL	1010	

DB	3114	LDQRTVTPDDEQKIAPDLTRS-AOP	FSAEPTKRKTVVGASDQAKASE-----	3162		
QY	1991	HATITARTPAKNLAPHAS--PDP	P-----APPASDPHREKTQSK	PFISIOLEL	2038	
DB	3163	--TIVDREGMTAPYREKSLCPD	TQHARNVLEWSPSDNLT-----	NPPAQGE--	3209	
QY	2039	RSGLYHGS-----SYSPG	VEVPVSPVSSPSLTUD-----	KGLPKH	2073	
DB	3210	-----HGSFSEGSTTDSRDAIT	DESATTSKGSEKVLDAISILAT	HDADHPPTVRNLPEH	3264	
QY	2074	LEELDKHLEGLRPKQPGVKLG	GEAAHLPH--LRPLPESQSSP	LLQTAGVKGHOR	2131	
DB	3265	ETQAPAS-----TQAS	VEASDEKAKSTDQEDMKP-----	ASQASILTQGTGDK-	3311	
QY	2132	VVTLAOHISEVITQDTRHH	QQLSAPLAPLAFYFGASC	PCFVLDLRRPPSDLYLP	PDHG	2191
DB	3312	--PIARFLHDQASPELTHPQ-----	-----PPES-----	PPHS	3340	
QY	2192	APARGSPHSEGGKRSPEPNKTS	VLGGGEDGIEPVSPPEGTE	PGCHSRSAVYPLLYRDGEQ	2251	
DB	3341	ASQHDAP-----TDVLDK	TMTKLTSTNQEGMAPTAGSR	STL-----	3377	
QY	2252	TEPSRMGSGPNTSOPPAFTSKLT--	ESNSAMVKSQKQEKINKL	ANTHNRNEPVNIS-	2307	
DB	3378	-----ESQPDGTFAAEV	VDHQKSTLSDQESARATQ	PLSVRSKEDTNVSA	3424	
QY	2308	--QPCT--EIFNMAITQGLM	TVRSQAVQEHASTNMGLEAI	IRKALMGKYDQWERSPP-	2362	
DB	3425	ADQEPVQTIQFHKARSSVPV--	TREVQFPDS-----	LGAISKIQEVS	PD	3467
QY	2363	--LSANAFNP-----IN	SASLPAAMPITADGRSD-----	HTLTPSGGGK	2402	
DB	3468	NQSTDYSAVQVPPTKEQVL	HAPQTSFDQEGITPAQGEK	SDIPDAQLVPHALQDKG--	3523	
QY	2403	AKVSGRPSRKAKSPAGLAS	DRPPSVSVSHSEGCNRRTP	L--TNRVWEDRESSAG--	2458	
DB	3524	-----SPRSTQPPPIESEKE--	SQTAEDG-----QIVLLQSS	VVQVSTESTPSLAGPR	3568	
QY	2459	-----STPFFYNPL-----	IMRLQAGMASPPPPGLP	PAGSGLAGPHH	2496	
DB	3569	ESSSDSPYSAENQVSTPKEVE	RLPPTTGMDRVQP---SSEAT	PGVVGISATPG---	3622	
QY	2497	ANDEPKP 2504				
DB	3623	--DNQSP 3628				

RESULT 46
Q9UH8 PRELIMINARY; PRT; 2296 AA.
AC Q9UH8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Splicing coactivator subunit SRM300.
GN SRM300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20132239; PubMed=10668804;
RA Blencowe B.J., Bauron G., Eldridge A.G., Issner R., Nickerson J.A.,
RA Rosonina E., Sharp P.A.;
RT "The SRM160/300 splicing coactivator subunits."
RL RNA 6:111-120(2000).
DR ENBL; AF201422; AAF21439.1; -.
SQ SEQUENCE 2296 AA; 251964 MW; 17C0BD4EA10A9CF9 CRC64;
Query Match 3.6%; Score 478.5; DB 4; Length 2296;
Best Local Similarity 19.5%; Pred. No. 1.8e-14;

QY	779	P----	PGPPTPRRTSRAP	IEPTPASEATGAPT	PPPPAPSPSAPP	PPVVPKEEKEETAAP	835										
Db	807	PAQFSSPP	PHGRSAHSLQ	FLVRQ-----	FNQVPEIL	UTEPRDPTEPEP	855										
QY	836	PVEEGEQK	PPAAEELAVD	TGKAEEPVKS	ECTEB--	AEEGPAKGK	890										
Db	856	PKPEPEK-----	TEEFQWPORT	QTLAQLPAEKL	PPKKRLRLAEMAQ	SGESS	903										
QY	891	LKAKEGGSGR	ATTAKSGAPO	SSSATS	CSADEVDEAEGG	KNRLLSRP	942										
Db	904	FESSVP--	LSRSPQSS	ISLSGSSRSAS	FDREDHGKAE-----	APGPFSDTRSKT	952										
QY	943	-----	SLITPTGDP-----	RANASPKP-----	LDLKOLKORAA	AIPIQV	978										
Db	953	LGSHMLTV	PHSHPHAREMR	RSASQSPNV	PHSHMTETRSK	SDYGLSTGFSL-----	1007										
QY	979	TKVHEPPRE	DAAPT	KPAPPAPP	PPQ-----	NLOPESDAPOP-----	G 1016										
Db	1008	-----	AVPAAPP	PAAPPERRK	CFVLROASLNR	PEAELEAVPREN	KRAVRMPAA 1057										
QY	1017	SSPRGKSR	SP-----	APPADKEAFA	ABAOKLP--	GDPPCWT	-----GLFPFV 1057										
Db	1058	SKPSTKSV	PQISVGT	TGGPSGK-----	SQMQRPP	PLGSSPPY	TEALQVFP	PLGTQLPP 1113									
QY	1058	P-----	PREVIKAS	HAPDPS-----	ARSYAPP	GHPLPLGL	HDHTAR	PVLPRTTI 1102									
Db	1114	PASLFS	LOQLLP	QQQSSFF	FTQAWAGLL	SSSYNP--	PLPPLFQA--	PPLPLOTV 1169									
QY	1103	SNP-----	PPLISSAKH-----	PSVILERO	IGAISQ	MSVOL-----	HV 1135										
Db	1170	LHPSQL	HLPLLPHAA	DIFFQPP	PSFLPMPCAP	STLSGVFLP	QSQFALQ	LQCEIESHL 1229									
QY	1136	PYSEHAKA	PVPTWGL	PLMDPK	LAPFG-----	VKQEQLSP	RCQAGP	BESLGVPTA 1189									
Db	1230	P---	PVKTS	PLPLATG-----	PPGSSSTE	YSDDIQLP	PVPTQATS	PAPTSAPP--	LALPAC 1281								
QY	1190	QEASV-----	LRGTAL	GSV-----	PGGSIT	KIGIP	STRVPS	DAITYRG 1227									
Db	1282	PDAW	VLVVR	VIQTHMPS	YCSAMV	TLSQILV	TQSPG	PASTAL	TKYEPSSK	MTV-- 1339							
QY	1228	SITHGT	PADVL	YKGTIT	RIIGED	SPRLDR	GREDS	LPGKHY	YEGKHVL	SYEGGMSVT 1287							
Db	1340	-----	CEADV-YE-----	AEPG	SSISKE	QNRGYQ	TPYLRV	PERK	GTLS	SSGILSLE 1386							
QY	1288	QCS-----	-----	KEDGR	SSSGP	HPH-----	-----	TAAPK 1308									
Db	1387	GCS	TASG	KRVLP	AGSL	MTMETQ	QOKRV	KEBEAS	KADEKLE	LVSTCV	SVLTS	EDRK 1446					
QY	1309	RTYD	MEGR	VGRA-----	ISSAS	TEGLM	GRAIP	PERH	SPHKL	EQHHIR	GSITQ	IPR 1361					
Db	1447	KTEK	PHVGG	QGRSR	REAE	TLS	SDVSD	PKEL	SPLSHS-----	-----	TL	SHGTAPG 1494					
QY	1362	SYVEA	QDYL	RRREAK	LKREG	TPPP	PPPS	DLT	BRAYK	TQALG	PLKPA	HGLVAT	KBA 1421				
Db	1495	S--	EALKEA	QPS	SSK-AHRR	GLPP-----	-----	MSVRK-- 1520									
QY	1422	GRS	THEI	PREEL	RHTPEL	PLAPRL	KEGSI	TQGT	PLKYD	TGAST	TGSKKH	DVRS	LIGSPG 1481				
Db	1521	-----	EDPKE	QDLP	PLAPP	SSLP	USD-----	TSRPAK	LOEQ-----	-----	TD	SKK--	V 1559				
QY	1482	RTFP	VPVH-----	PLDYM	-ADAR	ALERA--	CYEES	LKSR	FGT	ATSS	GGG	STAR 1524					
Db	1560	LQFP	SLHT	TNV	WCVL	YNV	IKPHN	HIQ	HADR	SSV	VYAG	WCISL	YNP	NLPG	VSTKA	ALSLLR 1619	
QY	1525	GAP	IV	PEL	GR	QRP	QPL	YED	HGAP	FACH	LPRG	SPV	TRE	TP	LOES	LSS	KASQDRK 1584
Db	1620	-----	SKQV	SKETY	TWTAT	AP-----	HP	EAG	LV	PSN	RKPR	MT	VEHL	PSV	SPESQK 1667		
QY	1585	LTST	PRE	IAK	SPHS-----	TVPEH	PHPT	SPYEH	LLRG	VSG	VDYLR	SHI	PLAF-----	1632			
Db	1668	----	DPA	VEKE	KQK	AREG	PTSK	RGE	PAR	KIF	E-----	GG----	YKS	NEE	YI	YVRGR 1715	
QY	1633	-----	DPTS	I	PR	G	I	PLD	AAAY	YL	PRHL-----	APN	TY	PH	LY	PPYL	IRG 1672

Db	1716	GRGRYVCE	GIRCKP	SMKKH	IRTH	TRTD	VRPYVCK	CHFAFK	TGNLT	KHMKSK	KAHKK 1773																																												
QY	1673	YPTD	AAENR	QTI	INDY	ITS	QOMH	HTATAMA	QRA	MDMLR	GLSPRESS	LALNYA	AGPRGII 1732																																										
Db	1776	COET	GVLE	BELEA--	-----	BEQTS	DDH	QD-----	-----	SEGQEG-- 1804																																													
QY	1733	DL	SQVPHL	PVLV	PPPT	PGTP	ATAM	DRLAY	PTAP	OPFS	SRHSS	SPSP	GGTHLT	TKPTTTS 1792																																									
Db	1805	-----	-----	AEAVE-----	-----	EHQF	SDLE-----	-----	-----	DS 1819																																													
QY	1793	SSER	DRDR	DRDR	REK	REK	SIL	TSIT	TV	VEHAP	TRP	CTG	SSSSSS	GGGSSSRPA 1852																																									
Db	1820	DS	DL	DL	DE	DE	EE	EE	EE	EE	DE	LS	GCPS	-EAP	PCLP	PT	LOENSSP----- 1864																																						
QY	1853	SH	SHA	HQ	SP	IS	P--	R	T	D	A	L	O	R	P	S	V	L	H	N	T	G	M	K	I	I	T	A	V	E	P	S	K	T	P	L	R	S	T	S	T	S	S	P	V	P	R	P	A	A 1911					
Db	1865	-----	-----	VEG	P	Q	A	P	D	S	T	S	DE	VEG	S	I-----	-----	SEATH	L	T	A	S	S	C	T	P	S	R	G	T 1903																									
QY	1912	T	P	P	P	A	T	H	C	P	L	G	T	L	D	G	V	Y	T	T	L	M	E	P	V	L	P	K	E	A	R	V	R	P	R	P	A	D	T	G	H	A	F	L	A	K	P	P	A	R	S	G	L	E	P 1971
Db	1904	GL	P	R	L	G	L	A	P	L-----	-----	E	K	D	M	S	S	A	P	S	K	A	T	S	P	R	P	M	S-----	-----	PS	KE	A	S	R	P 1942																			
QY	1972	A-----	SS	P	S	K	S	E	R	P	L	V	P	P	V	S	G	H	A	T	I	A	R	T	A	P	K	N	L	A	P	H	A	S	P	D	P	P	A	P	A	S	A	S	D	P	H	R	E	K	T 2026				
Db	1943	S	V	T	R	K	H	S	L	T	K	N	D	S	S	Q	Q	C	S	P	A	R	E	A	Q	A	S	V	I	T	S	T	P	G	Q	M	G	P	R-----	-----	D	L	G	P	H	L	C	S	P	R	E	L	S 1997		
QY	2027	Q	S	K	P	F	S	I	O	E	L	R	S	L	S	I	G	H	S	S	P	E	G	V	P	S	P	S	L	T	H	D	K	G	L	P	K	H	L	E	L	D	K	S	H	L	G	E	L 2086						
Db	1998	R	L	T	P	Y	P	I-----	-----	G	R	E	A	P	A	G	L	E	R	A	T-----	-----	D	T	G	T	P	R	S	P	T	R	S	L	G	O	A	E	S 2036																
QY	2087	R	P	K	Q-----	P	G	V	K	L	G	E	A	N	L	P	H-----	-----	L	R	P	L	E	S	O	S	S	P	L	O	T	A	P	G	V	K	H	O	R	V	T	L	A	Q	H 2139										
Db	2037	P	Q	T	V	L	P	G	K	M	A	L	A	G	C	P	S	A	D	K	S	G	L	G	L	G	V	P	R	A-----	-----	L	I	Q	V	P--	L	P	H-----	-----	T	L	L	S	R	S 2084									
QY	2140	S	V	I	T	O	D	Y	T	R	H	H	P	Q	L	S	A	P	L	A	P	L	Y	S	F	P	C	A	S	C	P	V	L	D	R	R	P	S	D	L	Y	L	P	P	D	H	G--	A	P	A	R	G	S	P 2198	
Db	2085	P	E	T	C	S	A	M	R	K	T	E	S	R	S	P	S	A	G--	-----	-----	P	H	D	F-----	-----	H	G	H	L	P	S	R	S	E	E 2128																			
QY	2199	-----	H	S	E	G	K	R	S	P	E	N	T	S	V	L	G	-----	-----	G	E	D	I	E	P	V	S	P-----	-----	E	G	M	T	E	P	G	H	S	R 2238																
Db	2129	N	L	F	S	H	L	P	H	S	Q	L	S	A	P	C	P--	-----	L	I	P	I	G	G	I	Q	M	V	A	R	F	G	A	O	T	V	L	P	G	C	A	A	W	S	G	F	G	G	S	D	L 2186				
QY	2239	S	A	V	Y	P	L	L	Y	D	G	O	T	E	P	S	R	M	G--	-----	S	K	S	P	G	N	T	S	Q	P	P	A	P	F	S	K	L	T	E	S	N 2279														
Db	2187	T	G	A-----	-----	R	E	A	Q	R	S	R	S	P	T	E	S	P	S	A	S	V	P	V	A	K	S	K	F	T	L	S 2221																							
RESULT 49																																																							
ID	O43161	PRELIMINARY; PRT; 3851 AA.																																																					
AC	O43161;																																																						
DT	01-JUN-1998	(TEMBLrel. 06, Created)																																																					
DT	01-AUG-1999	(TEMBLrel. 11, Last sequence update)																																																					
DT	01-OCT-2003	(TEMBLrel. 25, Last annotation update)																																																					
DE	BASSON protein (Hypothetical protein KIAA0434) (Fragment).																																																						
GN	BSN OR KIAA0434.																																																						
OS	Homo sapiens (human).																																																						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;																																																						
OC	Mammalia; Theria; Primates; Catarrhini; Hominoidea; Homo.																																																						
OX	NCBI_TaxID=9606;																																																						
RN	[1]																																																						
RP	SEQUENCE FROM N.A.																																																						
RA	Winter C., Dieck S. T., Bockmann J., Boeckers T., Kaempf U.,																																																						
RA	Sanmarti-Vila L., Laignaese K., Altrock W., Stumm M., Soyke A.,																																																						
RA	Wleackner P., Garner C.C., Gundelfinger E.D.;																																																						
RT	The presynaptic cytomatrix protein Bassoon: Sequence and chromosomal																																																						
RT	localization of the human BSN gene.;																																																						
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.																																																						
RN	[2]																																																						
RP	SEQUENCE OF 2281-3851 FROM N.A.																																																						
RC	TISSUE=Brain; Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,																																																						
RA	Ishikawa K.,																																																						

RESULT 49

O43161	PRELIMINARY;	PRT;	3851	AA.
ID	O43161			
AC	O43161			
DT	01-JUN-1998	(T=EMBLrel. 06, Created)		
DT	01-AUG-1999	(T=EMBLrel. 11, Last sequence update)		
DE	01-OCT-2003	(T=EMBLrel. 25, Last annotation update)		
DE	BASSOON	protein (Hypothetical protein KIAA0434) (Fragment).		
GN	BSN OR KIAA0434.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Winter C., Dieck S.T., Bockmann J., Boeckers T., Kaempf U.,			
RA	Samarti-Vila L., Langnaese K., Altmann W., Stumm M., Soyke A.,			
RA	Wiacaker P., Garner C.C., Gundelfinger E.D.;			
RT	"The presynaptic cytomatrix protein Bassoon: Sequence and chromosomal			
RT	localization of the human BSN gene."			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 2281-3851 FROM N.A.			
PC	TISSUE=Brain;			
RA	Ishikawa K., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,			

RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y18448; CAA77176.1; -
 DR EMBL; Y18449; CAA77176.1; JOINED.
 DR EMBL; Y18450; CAA77176.1; JOINED.
 DR EMBL; Y18451; CAA77176.1; JOINED.
 DR EMBL; AB007894; BAA23707.1; -
 DR PIR; T00062; T00062.
 DR GO; GO:0045202; C:synaptic junction; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR InterPro; IPR008899; Znf_piccolo.
 DR Pfam; PF05715; Zf_piccolo; 2.
 KW Hypothetical protein.
 FT NON TER 1
 SQ SEQUENCE 3851 AA; 410202 MW; 7CA91DB23284C161 CRC64;

Query Match 3.6%; Score 476.5; DB 4; Length 3851;
 Best Local Similarity 18.6%; Pred. No. 4.4e-14;
 Matches 634; Conservative 373; Mismatches 1163; Indels 1231; Gaps 146;

Qy 7 LVAQTWRATE-----PRYPHSLSYVQIARTHTDVGLEYQHHSRDYASHLSPGSII 59
 Db 438 LNCOTKELGSLGEPTLPPTSQPPVGNPHRASGTSPLKQK-----GPOGLG 487
 Qy 60 QPQRRRLSLSEFO-----PGNERSQELHLRPPESHLYLPGLKSEMEFIESKRRLLEL 113
 Db 488 QPSGPLPAKASPLSTKASPLPSKASPOAKPLRASEPSKTP-----SSVQEKTKRVPTK 540
 Qy 114 PDPLLRSPILLATQCPAGSEDLTKORSITGKLEVPSP-----PSPHTDLELELV----- 163
 Db 541 AEPMPKPPP-ETTPTPA-----TPKVGSGVRAEPATPVVKAVPBAPKGEAEDLVGKPPS 595
 Qy 164 -----PPRLSKEELION-MDRVDRITWVEQIISKLKKQOOLEEBAKPPPEKPVSP 216
 Db 596 QDASRSFQSLSDTCYSDGTSQSSQSEITGVVQ-----EVEQLDSAGVTGPHPPSPSEI 649
 Qy 217 PPIESKHSRLVQIYIDENRKAAEAHRLBGLGQVE-----LPLYNQPSDTRQYHENIK 271
 Db 650 HKVGSSMRLLQA-----QGLAPRSERKPLSSGTGEEQQRPHSLISITPEAFDSDELEDI- 705
 Qy 272 INQMRKLLILYFKRRNHARKQWKQKFCQ-----RYDQLEMALEKKVERIE 317
 Db 706 -----LEEDEDSEAWRRREQQDQTAESSDDFGSLRHVDYVESSEGLSPPLP 752
 Qy 318 NNPRRAKESKVRYYEKQPEIKQRELOERMOSRVGQRG-----SGLSMAAARSEHEVSE 374
 Db 753 PQPARAAELTDEDFMRQQLIEMSAEDNLEEDTATSGRLAKHGKQKGRPRRPFQ 812
 Qy 375 IIDGLSE-----QENLEKQMRQLAVIPMLYDADQQRKFINNM----- 413
 Db 813 EPAALPKRRLPHNATTGYEELLPEGGAETDGGTGLQGLRRFKTIELNSTGSGYGHED 872
 Qy 414 -GLMADPMKYKROVNMWSEKEKETREKFMQHP--KNFGLIASFLERTVAECVLYY 470
 Db 873 LGQGPDPF-----LDREPELEMESLTGSPEDSRGHSSTLPASTPS-----Y 915
 Qy 471 YLTCKENYKSLVRSYRRGKSQQOQQOQQOQQOQQO-----PMRSSOEKDEKEKE 526
 Db 916 TSGTSPISLSLESDSSFSRRORLEAKQQRKARHRSGLPPLPTIEDSEBELREE 975
 Qy 527 KEAKEKEPEVENDKDLKEKTDSDSGEDNDKEAVASKGRKTANSQGRKGRITRSM 586
 Db 976 ELLREQEKREVE-----QQRIRSTARKTRDKELRAQ-----RRERSKTPP 1019
 Qy 587 ANEANSSEAITP-----QOSALASMELNWSRWTEEMETAKKGLLEHGRNWSAIRMVG 642
 Db 1020 SNLSPIEDA-SPTBELRQAEMEBELHRSSEYSPSLDSEALDGGP-----SRL-- 1071
 Qy 643 SKTVSQCKNFYFNK--KRONLEILLQHKLEKERNARRKKKAPAAASEAAFPVV 700
 Db 1072 -----YKSGSEYNLPTFMISLYPTETPPSGSSTTPSGRPLKSAABA----- 1112
 Qy 701 EDEMEASGVSGNEEMVEBAEAL-----HAGSGNEVPRG-----ECSGPATVNNSSDTSIP 752

Db 1113 -----YEMMRKAELLQROQGAAGARGPHGSPQTPGRGLGSFHYQDITD 1159
 Qy 753 SPHTAAKDTGONGPKPPATLGA----- 775
 Db 1160 REYQQAQAQPAEG---TPASLGAAYEEILOTOSIVMRQOASSRDIAFADKKKEQFL 1216
 Qy 776 -----DQPPPGPTTPRRTSRAPIE---PTPASEATGAPT- 808
 Db 1217 NAEAYMDPMKQGGPLTPTGTSPTQLA-APVSFTPTTSSDSSGGRVTPDVRVTOHFAKET 1275
 Qy 809 -----PPAPPS-----PSAPPVVPKEKEEETAAAP-----PVEGE 841
 Db 1276 QDPLKLHSSPASPSASKEIGMPFSQGPGRP-----ATTAVAPCAGLPRGMYMTPASAG 1330
 Qy 842 EOPPPAAEELAVDTGKAEPEPVKSECTEAEERP-AKGDKEAAEAATAEGALKAEKKEGS 900
 Db 1331 SERSPSSSTAHYVG--HSPITANYGQTEDLPQAPSLGAAGRAAREKPLASDGEGET 1388
 Qy 901 ---GRATTAKSSGAPQDSDSS-----ATCSADEVDEAEGDKNRLL-S 939
 Db 1389 PQPSRAVSYFASSPPLSPSSPSESPTFSPCKMGPRATAEFSTQTPAPASDMPRSPGA 1448
 Qy 940 PRPSLLTPTGDPANASPOKPLDLKQLKQRAAIPPI-----QVTKVHEPPRED 988
 Db 1449 PTPSPMVAQG---TOTPHRPSTPRLVWQESSQAEPMVITLASDASSQTRMVASASTS 1504
 Qy 989 --AAPTCKPAP-----PAPPQNQLQPEDAPQOQSSPRGKSRSPAPPADKE----- 1033
 Db 1505 PLCSPTSTQPTTHGYSTQTPSPVSQLPPE-----PPG--PPGPRVPSAGADPLALYGM 1557
 Qy 1034 -AFAAAQKLPQDPPCWTGSLPFPVPPREVIKASHAPADPSAFSAYAPGHPPLGLLHD-- 1090
 Db 1558 GALPAENISL-----CHISS-----VPGTSRVEPCPTPTGTAVDLRTAVKPTLITDQG 1608
 Qy 1091 ---TARVLPBRPTIS-NP-PPLISSAKHPSVLE---RQTGAISQGSVQLHVYSEHAK 1142
 Db 1609 MDLTSLAVERKYGALDPIPGRQSTAVQPLVINLNAQEHFTLATATTVSITMASSVFMA 1668
 Qy 1143 APVGPVTMGLPLPM-----DPKKLAPFSVKQE-----QLSPRGQAGRPES----- 1183
 Db 1669 QQQPVVYGVDPYQSRDLDFGQGGSPVCLAQVKQEQAVQIAPYRSGRGR--PREKPAR 1726
 Qy 1184 LGVPTAQEASVLRTALGSPVGGISITKGIPTSRVPSDSAITYRGSITHGTADVL----- 1238
 Db 1727 YNLFN-QVAPLARDVLITQWGTAAQSIGLXPGVPPEGAEPHR-----ATPAELRSHALP 1780
 Qy 1239 -----YKGTITRIIGEDSPSRLD---RGRBDSLPGKHVIEGKKGHVLSY 1280
 Db 1781 GARKPHTVVVQMGEGTAGTITLLPEEPAGALDITGMRPESQACDMMVKLPFGS--SC 1838
 Qy 1281 EGGMSVTQCSKEDGRSSSGPPHETAAP--KRTYDMMEGRVGRAISSASIEGLMGRALPP 1337
 Db 1839 TGFTHPAPSVEKSMADAPPQSSSPFYGPRDPEPPEPTTYRA-----QGVVGPQ-PH 1891
 Qy 1338 ERHSPPHLKEQHHRISITQIGIPRSYVQAEQEDYLRREAKLLKREGTTPPPPPSRDLTEAY 1397
 Db 1892 EEQRPY-----PQGLPGLYSSSDTNLAAGL-----NY 1921
 Qy 1398 KTOALGPKLKPAAHEGLV-----ATVKEAGRSIHEIPRELRHTPBLPLAPR 1444
 Db 1922 HAQRIGQLFQPGRGSAMDLSLKHYSYLGADRYLGQGLQYGVSDLRHPTDLAHL 1981
 Qy 1445 PLK-----EGSITO----- 1453
 Db 1982 PMRRYSVSNYIHDRHYRPGDVGFOEASLAQYATTAREISRMCAALNSMDQYGRHG 2041
 Qy 1454 ---GTP-----LKYDTGASTTGKXHDVRSLLGSPGRTTPPVHP----- 1489
 Db 2042 SGGGPDVLVQYQOHGPGLSAPQSLVPLRPLCLGNP--TFPEGHPSGNLAQYCPAAGQ 2099
 Qy 1490 -----LDVNADARALERACYEE-----SLKSRPGTA-----SSSGG-- 1520

D	b	2100.	TAVRQLLPSTATVRAADGMITYSTINTPIAATLPIITQPASVLBRPMVRGGMRYRPAAGGIT	2159
Q	y	1521	-----SIARGAPVIVP-----LCKP-----	1536
D	b	2160	AVPLTSLTR-VPMIAPRVPLGPTGLYRPAPSRFPPIASSVPPAEGFVYLKGKPAAAKAPCA	2218
Q	y	1537	-----RQSP-----TYEHG-----APFAGH-----LPRGSPVTVREPT	1566
D	b	2219	GGPSRPMVPGAAREEPLPTTTTAAIKEAAGAPAPLAGOKPPADAAEGGSGALLSREG	2278
Q	y	1567	PRLOEGSLSSKASQDKLSTPRE-----IAKSPH-----	1597
D	b	2279	FEKEEAQEEBRQKQEQLLQLERERVELBEKLRQLRLOEELERVELQHRHEEEQLLVQ	2338
Q	y	1598	-----STVPEH-----HPIPSYPYHLLRGVSGVDLYRS	1626
D	b	2339	RELQELQTIKHHVILQOOQEEBQAFQALORLEQIQLOOQLOQOOLEEQOKQ	2398
Q	y	1627	HIPAFDPTSIPIRGIPLDAAAAYVLPRIAPNTY-PHL-----YPPY	1668
D	b	2399	KAPFPAAACEAFGRGPPAAAA-----ELAQOQYWPPLTHAAFIAWAGPEGLQGPREPV	2451
Q	y	1669	LIRGVPDTAALENQOT-----II	1686
D	b	2452	LHRGLPSASDMSLQTEOWEASBSGKIKRHSMPRLRDACELESGETPCVVVRIADSSVQ	2511
Q	y	1687	ND-----YITSOO-----MMH	1697
D	b	2512	TDDEGSESYLLSRRRRARRSADCSVQTDDEDAEWEOQPVRRRRSRLPRHSDSGSDSKHD	2571
Q	y	1698	NTATTAQAORA-----DMLRGLSP-----	1715
D	b	2572	ATASSGAAATVRAMSVGIQTIIDSCSVQTEPDQLPRVSPAIIHTAATDPKVEIVRYISA	2631
Q	y	1716	-----RESSALNY-----AAGPRGIIIDSOVPHLPVLVPPPTPGTATAMDR	1757
D	b	2632	PEKTGRGESLACQTEPDQOAGVAGPOLVGTALSPYLPQIIVTPG-PLGRFEKKKPKDP	2690
Q	y	1758	-----LAYLPTAQPPSSHSSPSLPGGPTHLTKTPTTSSSERERDRDR	1804
D	b	2691	LEIGVQAHLPPESSQLSVRQPPKSPQVLYSPVPSPLSP-HRLDTSFASER-----L	2742
Q	y	1805	DRDREREKSIIITSTTTTVEHAPIWR-GTEQSSGSGSGSGGGGGSSRPASHSHAHQHSPI	1863
D	b	2743	NKAHVSQPKHFTADSALRQQTLPREMKTLQRSLS-----DPKPL	2781
Q	y	1864	SPRTQDALQORPSVL-HNTGMKGIIITAVEPSKPTVLSTSSSPVRPAATFPATHCPLG	1922
D	b	2782	SPTABESAKERFSLYQHOGGLGQSVSALPPN-SLVRKVRKTLPSPP-----PEEAHLPLA	2835
Q	y	1923	G-----TLDGVVPTLMEPV-----LILPKEAPVAR-----PERP	1951
D	b	2836	QOASPOLYAASLLQBLGTGPTTVATKXALLRELDRLRLVEHSTKLKKQAELEDDEEK	2895
Q	y	1952	RADTGHAFF-----LAK-----PPAR-----	1966
D	b	2896	EIDAKLKYLELGITQRKESLADGGRDYPPLRGLGEHRDYISDSLNQLRLOQCTTPAG	2955
Q	y	1967	-----SGLBPA-----SPSKGSERPVLPVVSGHATITARTPAKNLAPHASDPDPAP	2014
D	b	2956	QVDFPPTAAAPATPSGTATFOQPR-FQPPAQVYSAAGSGGTQNGFPAHQAPTYPGSTY	3014
Q	y	2015	-----PASASDPHREKTSQKFPFSIQELRLSLGYH-----SSYSPGVEPVSPVS	2060
D	b	3015	PAPAFPPCASYP-----AEPGLPNQQAFRPTGHVAGQTPWPTTQSTLFPVPADSRAPLQ	3068
Q	y	2061	SPSLTH-----DKGLPKHLEELDKSHELELPKO-----PGP-VKLGGEAHLPHL-----	2106
D	b	3069	KPROTSLADLEQKVTNTEVITASPVVPMSSAPSETSYSGPAVSGYQGVKVPVPRAGDR	3128
Q	y	2107	-----RPLPESOPSSS-----PLIQATAPGVKGHQRVVVTLAQHITSEVITQDTRHHP	2157
D	b	3129	GSVCSOPAP-TYPSDSHVTSLFONVP-----BNVMTIDDIJSE-LTKDSTSTAPDSOKLE	3180

Qy	2158	PLPAPLYSFPGACSVLDRRRPSDYLPPDHGA PARGSPHSEGGKRSPEPNKTSVLGG	2211
Dd	3181	PLG-----PGSS-----GRPKEPGEPPGVLDG	3202
Qy	2218	-----CEDGLEPV-S-PPEG-----WTEPHGSHRSAYV--PLLRYRDEOTE---PS	2255
Dd	3203	FTU.PCCVARGEESEEDSYDPKGHLRMENSGRPASTHYVGDSYRGARVEKYGPG	3262
Qy	2256	RMGSKSPGNTSQPAFESKLTESNSAMVKSKQEIINKLANHNREPEYNISOGTFEIN	2315
Dd	3263	PMGPKEPSKSIAAATSSKES-----KHKKQGMEQIKSKFGPIEEKADVE---SDLAS	3312
Qy	2316	MPAITGTGLMTYRSQAQVEHA STMGLLEALIRKALMGKYDOWEESPPLSANAF--NFLNA	2373
Dd	3313	YPFPAVVSSSLVSRRKFQDEIT-YGLUKXN-----YEQQKYGMSSDAVEDDRIYG	3363
Qy	2374	SASLPAAAMPITAADGRSDHTLTSPGGGKAKVSGRPSSRKAKSPAGLAGSDRPPSVSV	2433
Dd	3364	GSSRSRAPSAVSGEKLSHDIFS---GWGKGYEREREAVERLQK-----AGPFSSLSMA	3414
Qy	2434	HSEGDCNRRTPLTNRWEDRPSSAGSTPFYNNPLIMLRQAGVMASPPPPP	2483
Dd	3415	HS-----RVAPPMSQAASEESPVSPLGRPRPAG---GPLPPGGDTCPQCFS	3459
Qy	2484	--LP-----AGSGPLAGPHANDEEPKLLCSQYETLSDSE	2517
Dd	3460	HSMPDQVGHVKDGPR---HAYKREEGYILDDSHCVVSDSE	3497
RESULT 50			
Q9UPA5	ID	PRELIMINARY;	PRT; 3926 AA.
AC	Q9UPA5;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Neuronal double zinc finger protein.		
GN	ZNF231.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBT_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93026125; PubMed=9806829;		
RA	Hashida H., Goto J., Zhao N., Takahashi N., Hirai M., Kanazawa I.,		
RA	Sakaki Y.;		
RT	"Cloning and mapping of ZNF231, a novel brain-specific gene encoding		
RT	neuronal double zinc finger protein whose expression is enhanced in a		
RT	neurodegenerative disorder, multiple system atrophy.";		
RL	Genomics 54:50-58(1998).		
DR	EMBL; AF052224; AAC83555.1; -.		
DR	Genev; HGNC:1117; BSN.		
DR	GO; GO:0005634; C:nucleus; TAS.		
DR	GO; GO:0007268; P:synaptic transmission; TAS.		
DR	InterPro; IPRO08899; Znf.piccolo.		
DR	Pfam; PF05715; Zf.piccolo; 2.		
SQ	SEQUENCE 3926 AA; 416458 MW; 75CD56EADD437343 CRC64;		
Query Match			
Best Local Similarity 18.6%; Pred. No. 4.5e-14;			
Matches 634; Conservative 373; Mismatches 1163; Indels 1231; Gaps 146			
Qy	7	LVAOTWRATE-----PRYPHSLSYVPQIARTHTDVGLLEYQHHRDYASHLSPGSII	59
Dd	513	LNCQTKRLLGSLGETPLTPPTSQPPVPGAPHRASGTSPLKQK-----GPOQLG	562
Qy	60	QPQRRLPSLLSEFO-----PONERSQEIHLPESHVYLPGLCKSEMFISKRRPLELL	113
Dd	563	QPSGGLPAKASPLSTKASFLPSFKASPQAKPLRASFSKTP-----SSVQEKTKRVPTK	615
Qy	114	PDPIARPSPILATGAPAGSEDILTQRSITLGKLEVPSP-----PSPPHTDPELEV----	163

Db 1033 GRNLGHIYSDSESSSEKDEOQATVESNVDSQNOQTIRTKAAMKEFVPGTAATTS 1092
QY 908 SSCAPQDSSTATCSADEYDEAEGDKRLLSPRSLTPTGDPANASPOKPLDLKOLK 967
Db 1093 STSQASSTSKAKNTRGEGGAIGNSTKTKPN-----AKLYPADLLVVP 1140
QY 968 QRAAAIPIQVTKVHEPPREDAAPTKAPAPPQNPQLOPE-SDAPQOP-----GSSPRG 1021
Db 1141 ORQAA-----KKASENRSTNLAT-----TLQPDVSDVRBPDIINSIGTAKS 1183
QY 1022 KSRSPAPADKEAFAAEAQKPGDPCWTSGLPFVPPRE-----VIKASPHAPDPSAFS 1076
Db 1184 KYKDSRSSRVNSNEADKSSLEK-----VRPKEHLQKTVGKTSAPAEGRK 1228
QY 1077 YAPPGHPLPLGLHDTARVLPRLPPPTIS-NPPPLI---SSAKHPSVLERQIGAISQMSVO 1132
Db 1229 GRPPKVPK-----ARPPSITENDKALPTHQSKPPSVVATPSAKS-NFAVS 1276
QY 1133 LHVYSEHA-----APVGPVTMGLPLPMDPKKLA 1162
Db 1277 L-VPQRAAKAAEQKSKPVLESFSTGNDISDKETVTSATISGSGSVPAASTPVKPT 1335
QY 1163 PFSGVKQOLSPR-----QOAGPPESL-----GVPT-----AGEASVL 1195
Db 1336 RRSIKEAPITPKELSGRRKSKEDLLATPIKTTPLVRRVVVNLSSSGSDSSSSS 1395
QY 1196 KGTALGVPGGSITKG-----IPSTR-----VPSDAITYRGSITHGTADVLV 1239
Db 1396 SSSGSSSSGSDSSESQASNSNPSSREPPVAPAKVPSDSSIVPKKS-----PRKSM- 1449
QY 1240 KGTITRIIGEDSPSL-----DRGREDSLPKGHIYEGKKGHVLSEYEGM- 1284
Db 1450 -----DKPSALTIAPASVWVNLTPSTRSQNSTKSKVALQKA-VQSVEDDVK 1497
QY 1285 -----SVTQSKSE-----DGRSSGPPHETA-APKTYDMGEGVRGAI 1322
Db 1498 CTPKTNRLQGSMECGKQVLEQATKTRATRGSKRPPSPPTAKSSPEKT-----VSRCK 1550
QY 1323 SSASIEGLMGRAIPPERHSPLHLK-EQHHRIGSITQIGIPRSYVEAQEDYLRREAKLLK 1380
Db 1551 SRA-----ESPKVANLEQEIQRKVASKGTSSLD-----KLNK 1587
QY 1381 -----EGTPPP-PPPSRDLTEAYKTQ-LGFLKLKPAHEGLVATK-----EA 1421
Db 1588 KQQQNMHSAQATPPISPTTASERIVKQCDLAKPDEVSIIQINLQADAPPEPLDPES 1647
QY 1422 GRSIHEIPR-----EELRHTP-----ELPLAPRLKEGSIQTGTPLKY 1459
Db 1648 AAEAGELPMDIDEELTTAPTQTLSASASKLADIIDDERPPAAPLPASPTPTTSNDMS 1707
QY 1460 DTGASTTCKKHVDVRLSIGSPGRTFPP-----VHPLDVADARALERACYESLSKR 1511
Db 1708 DAGSDLSERRRWRRRRRRRRRRHHPEHHTHTHQLINEMEMARELE-----EERKNEL 1763
QY 1512 PGTASSGSGSARGAPVIVP-----BL-----GKPRQSPLATVEDHGAFFA 1551
Db 1764 LANASKYASTSSPAVTVIPDPPEI IELDSNSAEQQQHLHDQPLPPLVVQSPAADV 1823
QY 1552 GH-----LPRGSPVTMREPTPL-----QEGSLSKASQDRKLTSTPREIAKSP 1596
Db 1824 PTVMQOQLLSPORPLIEQLPVEHLPIVETILEMEDSKFANFNASVLNPPNQMSLIG 1883
QY 1597 HSTVPEHHPHTISPYEHLRLGSGVDLYRSHIPLAFDPTSIPRGIPLDAAAAYLPRHLA 1656
Db 1884 SSI-----DRSKQISEDSIQATNLNLEKURKTKRKAQDDCCSKEAVDL-----LP 1929
QY 1657 PNPTPHLYPPYLRGYPDTALENRQTIINDYITSQOMHNTA-----TAMAQADMLRG 1712
Db 1930 TPAIPSVFP-----FHNADPEDIIHAQEQOHOQ 1960
QY 1713 LSPRESSIALNY--AAGPRGIIIDISQVPHLPVL-----VPTPT-GTPATA-MDLRA 1759

RESULT 52
Q9W0U2

Db 1760 Y---LPTAPQP-----FSSRH---SSSPLPGGPTHLTKPTTSS----- 1793
QY 2020 YQVNPNSQPPTLGLFLEKSPHQKACPLSSNGGANVGQAPPTDFVDLAAAANKVNTLG 2079
Db 1794 -----SERERDRD-----RERDRDRERKS----- 1813
QY 2080 SFRGAATVPTQSGTGVNAKINDYDESTRMQSPFGGMPWNESDLIAERRSSSPSSVESND 2139
Db 1814 -----ILTSITTVHEHPIWRPGTEQ---SSGSSSG-GGGSSSRPASH----- 1854
QY 2140 PPQPPVVVTTATARTARSQALESCKNFNFSPGNAGPCTAANATAPNHPMPVNGIDSI 2199
Db 1855 -----SHAHQSPISPRQTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSPVRP 1909
QY 2200 PMFNNTNTQOPTTFAHQOQOORTPNNOYN-----GTIYPOLAGIMHPOTT----- 2246
Db 1910 AATPPATHCPLGGTLDGVYPTLMBPVLLPKEAPRVARPERPRADTGHAFLA-KPPARSG 1968
QY 2247 -PTEPPSSLYGNGGVGAVQSTTLPP---PAQVNOY--PGTPYSATTLGMSVQOPALST 2300
Db 1969 LEPASSPSKSEPRPLVPVSGHATIAARTPAKNLAPH-----ASDPDPAPASADPHRE 2024
QY 2301 V-PVQTATPNPFTLTSPIDG--KMPTYPAQLLSSCAEAVVASMMPPTPVVAT---AK 2354
Db 2025 KTQSKPFSIQBLELRSLGSLYHSGSSYPEGV-----EPVSPVSSPSLTHDKGLPKHLEL 2077
QY 2355 DSPSKRTSVSSSNLSKKQTHKSPQLPOKSPGKSPRQLOPPTPPAPVVPVALPTKYDP 2414
Db 2078 DKSHLEGLRKPQ-----GPVKLGGEAAHLPHLRPLPESQSSSPL 2120
QY 2415 QTHTLQKGRQAPRAGSGSGGAPGRGRGRGRGGVTSQMAVMVLPMPMSDYGSNTHI 2474
Db 2121 -----CTAPGVKHQHVTVLAQHISEVITQDYTRHHQOQLSAPLAPL 2163
QY 2475 VNNLVGTPFEPNNEFDDMAGFVENLQSL-----RDRRRSFELRAP---RV 2517
Db 2164 YSFCASCPVLDLRRPPSDLYLP-PPDHGAPARG--SPHSEGGKSPBNKTSVLGGED 2220
QY 2518 QNKPTTPTTATTNPLHVPVLPFGVDMRTNVLGFEAPHSTA---SQEAYQNNLLGAFDS 2574
Db 2221 GIEPVSPEGMTPECHRSAYVPLYLRDGEQTEPSRMGS-----KSPGNTSQPPAFF 2272
QY 2575 GTADQTLSEFNEEDERQFOSA---LRATGTGTSPSKQHSQGTALVAPPTGNPTAPNLL 2631
Db 2273 SKLTESN-----SAMVKKKQEI-----NKKLNTHNRNE 2301
QY 2632 LHCTEANOAPNVAATGAATHLVEGSLVEASLEATSEVSDSDSTIPIHSTSTSDARSQ 2691
Db 2302 PEYNISOP-----GTEIFNMPAITGTGLMTYRQAOEHASTNMGLEAIIRKALMGKYD 2355
QY 2692 IKLKIKSPMAYPEHYNAMNTSSUTLSTLVQSSNVVQTTVSTVSA---SSAVSGNSR 2749
Db 2356 QWESPPLSANAFNPLN---ASASLPAAMPITAADG---RSDHTLITSPGGGKAKVSGRP 2409
QY 2750 RMRKELLSLVVQKDNHNDSSCGLPAASDTLPLENLKSEEDDELGGNGTKFKKNS 2809
Db 2410 SSRKAKSPAGLA-----SGD-RPDSVSVHSEGCNCRTPUTNRVW 2450
QY 2810 SSRELRDALNALVEQLLSSGACGGSGGRRRRSACSSGNSNNDNGKTGAASAG 2869
Db 2451 EDR-----PSSAGSTPPFPYNPLMRQAGVYASPPPPGLPAGSGPLAGPHAWDE--- 2500
QY 2870 KRRGRSKTLESSEDDHOAPKIKIR---GLUTANETPSGV---SSVDEGQNTSYEMTRRA 2923
Db 2501 -EPKPLLCQSYETLSDSE 2517
QY 2924 CPPKKRLTSNFTLTLLEE 2941

QY 1724 YAAGPRGIIDLSQVPHLPVLVPTTCTPATAMDRLAYLPTAQPFSSRRHSSPLSPGGPT 1783
DB 1654 YAS-----VVIIDHPEVTIEDP-----QSNLNDGPT 1682
QY 1784 HLTKTTTSSRRDRDRDREREKSIILSTTTTVEHAPIWRGTQSSGSSGG 1843
DB 1693 EV-----VSKQKRLQDEER-----QVNNKNAEGRS----- 1720
QY 1844 GCGSSRRPASHAHQHSPISPR-----TQDALQORPSPVHLNHTCMKGIITAVPSPKPTVLR 1899
DB 1721 -----QTSKLPREFAKQATGLOQAQS-----SASVPLASAPLPPTSASV 1762
QY 1900 STSTSPVRPAATFPATHCPGLTGLDGYVP-TLMEPVLLPKAEARVARPERPRADTGA 1958
DB 1763 PASTSAPL-----PATLTPVPASTSADVPASTLAPVLASTSAPVPASPLAPVASASV 1815
QY 1959 FLAKPAPS-CLEPASPSKSEPRPLVPPVSHATI-----ATPAKN 2001
DB 1816 SASVFASTSAAAITSSAPASAPATPILASVSTPASVTILASAPILASALASTSAPT 1875
QY 2002 LAPHASDPDP---APPASADPHREKTKQSKPFSIQELESLRGVHGSYSPEGVPVSP 2058
DB 1876 PAPAASSPAAPVITAPTIPASAP-----TASVPLA-----PASASAPAP-TP 1918
QY 2059 VSSPSLTHDKGLPKHLELDKSHLEGLRKPQGPVKGGEAAHPLHLRPLPESQPSGS- 2117
DB 1919 VSAPN-----PAPPAPAQTOAQ-THKPVQNPQLTTSQSSKQ 1953
QY 2118 -----PLLOTAPGVKHQVVTTLAHISEVITODYTRHHQO---LSAPI-----P 2160
DB 1954 PPPSIRLPSAQTPNG-----TDVASGKSIQTPOSHGTLTAEIWMNDKNVAPP 1999
QY 2161 APLYSFPGACSVLDLRRPPSDLYLPPDPHGAPAR-----GSPHSEGGKRSPEPNKTSV 2214
DB 2000 AVLNDISKLGPI-----SPPO-----PPSVANWNPILTSFGAPSEGAKNQOE-----SG 2046
QY 2215 LGGGEGIEPVSPPGMTPEPGHRSVAVPLLYRDGEQ--TEPSRMGSKSPGNTSQPPAFF 2272
DB 2047 LEIGDTTIQFAPASNGE-----NEVVPVLEKSAADKIPEKQORQOP---RAGPIKA 2098
QY 2273 SKLTSNSAMVSKQO-EINKKLANTHNN-----EPFYN----- 2305
DB 2099 QKLPDLSPVENKEHKEPGPIGKERSLKNRKVDAAQOQVEPEGOBKSPATVSTDPVTTKET 2158
QY 2306 ---ISQPGTEIFNMALITGGLMTYRSQAOEHASNTMGLEAITRKALMGKYDQWESPL 2363
DB 2159 KAVSESTEIGMISVSAEYGNKAKESVDTITPSSSLPNTVATNTKMETLVNNVPL 2218
QY 2364 SANAFNP-----LNASASLPAAMPITAADGRSDHTLTSFGGGGKAKVSGRPSRKAQSP-- 2417
DB 2219 PNTLPLPKRETIQSSSLTSVPPTT-----FSLTFKESARKAWENSPNVRKGSPT 2271
QY 2418 -----APGLASGDRPPSVSHSGDCNR-RTPLTNRWEDRPSAGSTPPPNPLMR 2470
DB 2272 STAPPIATGVSASSAGSPSTANTNSFSSAMPQIPVASVTPPTASLSCAGT-----YTTSSL 2327
QY 2471 LQAGVMSAPP-----PPCLPAGSGLA 2492
DB 2328 TKSTTTSDPPNICVKVPOOLOQTSLSLPSA 2355

RESULT 54
QSC0A3
ID Q9COA3 PRELIMINARY; PRT; 2219 AA.
AC Q9COA3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA1760 (Fragment).
GN KIAA1760.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
EX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
DR EMBL: AB051547; BAB21851.2; --
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR000719; Prot_kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR008271; Ser_thr_kinase.
DR ProDom: PD000001; Prot_kinase.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 2219 AA; 234470 MW; B7AD70196C2A92C0 CRC64;
Query Match 3.5%; Score 459.5; DB 4; Length 2219;
Best Local Similarity 20.1%; Pred. No. 1.5e-13;
Matches 564; Conservative 294; Mismatches 961; Indels 985; Gaps 146;
QY 55 PGSIIOPOP-RRPSLLSE-----PQGNERSQELHLRPEHSHYSLPELGKSEMEFIESKRP 108
DB 6 PGTLMPEGRGAGPAGMAEPRAKARPQPF-----LRRSVVESDQEEPP 50
QY 109 RLELL-----PDPLLR-----PSPLLATQOP-----AG 131
DB 51 GLEAAEAPGPQPQLQRRVLLCKTRLIAERARGPAAAPALVAQPGAPGAPADAG 110
QY 132 SEDLTOKRSLSGLKLEFVSPSPPHPTDPELELVPPRLSKEELIQNMDRVDRITWVEQQIS 191
DB 111 PEPV-----GTQEP--GPDPIAAAVETAPADGGPREAAATVRKEDEGAARKEPFG 161
QY 192 KLKKKQOOLEE-----AAKPPPEKPVSPPTIES 221
DB 162 RTRDEPEEEDDEDLKAVALSDGRFLKFDIELGRSGFTVYKGLDTETWVEVWVCEL 221
QY 222 KHRSLIVQIYDENRKAHAHRIIEGLGPQVELPLYNQPSDTRQYHENIKINQAMRKLI 281
DB 222 QDRKLTKL--ERQRFKEA--EMLKGL-----QHPNIVRFY--DFWESSAKGRCI 266
QY 282 L-----YFKRNHARKQWKQKFCQRYDQLMEAL-----EKKVER 315
DB 267 VLVTLMTSGTLTKYLRKFKVKKPKVLSWCR--QILKGLLFLHTRTPPIIHRDLKCDN 323
QY 316 I-ENPRRAKESV-----REYKQF---PEIRKQRELQERMQSVGQSGLSM-S 364
DB 324 IFITGPTGSKIGDLGLATLKRASFASVIGTPEFMPEMVEEHYDESVDVYAFGCMLE 383
QY 365 AARSEHEVSEIIDGLSEQENLEKQWQLAV-IPPMLYD-----ADQ 404
DB 384 MATSEYPYSEC-----QNAAQIYKVTGCGIKPASFEKVHDPEIKELIGECICKKEER 436
QY 405 QRIKFINNGLMADPMKV-----YKDRQVNMWSEQEKETFRKFKWQHPKNFGLI 454
DB 437 YEIKDILLSHAPFAEDTGVVRVLABEDHGKRTIALRLWVEDPK-----KLKGPKNDAI 491
QY 455 A-SP-LERKT---VAECVL---YYLTNKNENYKSLVRSYRRR-----GKSQ 494
DB 492 EFTFLEKTEPDEVAQEMIESGFPH-----ESDVKIVAKSRDRVALIQWRRERIWPAIQ 546
QY 495 QQQQQQQQQQQQQQQQMP-----RSSQEEKDEKEKEAEKEEKEPEVENDKEDL 545

Db 547 PKEQDVGSPDKARGPPVPLQVQVYHAQAQGPPEPE-EPEADQHLLPPTLPTSATSL 605
Qy 546 LKEDTDDTSGEDNEKEAVAGKRTANSQGRKRITRSMANEAN-----SEE 594
Db 606 ASDSTFD-SGQGS-----TVYSDSQSQSVMLGSLADAPSPAQVCVCPVSEB 654
Qy 595 AITPOQSAELASMELNESSRWTEEMETAKGGLLEHGRNWSAIAIARMVGSKTVSOCKNFYF 654
Db 655 PVLQSLSLGAY-----QOPTAAGL-----PVGSVFAPAC-----686
Qy 655 NYKQRQNLDEILOQHLKQKERNARRKKKAPAAASEAAFPVVEDEMEASGVSGNE 714
Db 687 -----PPSLQOH-----FP-----DPAMSFAPV-----704
Qy 715 EEMVEEABALHASGNEVPRGSCGPATVNNSSDTESTPSPTTEAAKQGTQNGPKPPATLG 774
Db 705 -----LPPST-----FMPTGPOQAPP-----G 723
Qy 775 ADGPPPGPPTP-PRTRAPTEP-----TPASEATGAP-----806
Db 724 QOPPLAQPTLPOVLAPQVVPVLPQVPPHPLPYLAPASQ-VGAPAKLPLQMPQAPLOP 782
Qy 807 ---TPPPAPPSPAP-----PPVVPKEKEEETAAAPVVEGE-----841
Db 783 LAQVPPQMPPIPVVPIITPLAGIDGLPALP-----DLPTATVPVPPPPQYFSPAVILPSL 838
Qy 842 -EQKPPAAEELAVDTGKAE-----PVKSECTEEAEGBAKDAEABATAEAGALKAEKK 896
Db 839 AAPLPFASPALPQAVKLPHPPGAPLAMPCTIIVNAP-----876
Qy 897 EGGSGRATTAKSGAQSDSATSACSADEVDEAEAGDKNRLSPRSLLTPTGDP-----951
Db 877 ---ATIPLLAVAPGVAISHTSA---VAQLFGQVYPAAFQ---MADTDVPPSPHH 925
Qy 952 ---RANASPKQLDLQLKQRAAALPIQVTKVHEPPREDAAPTKPAAPPAPPPQNTLOPE 1008
Db 926 TVQNMRATPPQ-----ALPP-----QPTLPPQVPLPPQPTLPPQ 960
Qy 1009 SDAQOQSSPRGKRSRAPPADKAEFAAEQAQKLPDPPCWTSGLPVPPVPREVILKASPH 1068
Db 961 PVLPPQ-----TRPQVPLPPQMLPPQVPLPPQ-----ALVREPL-----QPH 1003
Qy 1069 APDSAFSAPP-----CHPLPLGLHDTAR-PVLPAPP--TISNPPPLISSAKHPSVLER 1120
Db 1004 LPEQAAPATPGSQILLGHAPYADVAAQVTVVPVPAAVLSPLPEVLPAPAPELLPO 1063
Qy 1121 QIGAIQ-GMSVQLHVPYSEHAKAPVGPVTWGLPLPMDPKKLAP-----FSGVKQROL 1172
Db 1064 FPSSLATVSASVQSVPTQTATLLPPANP-----PLPGGPGIASPCTVQLTVEPVQEBQA 1118
Qy 1173 SPRQAGPPESLGVPTAQEASVLRGTALGSVPGGSITKGISTRVPSDSAI-TYRGSITH 1231
Db 1119 S---QDKPP---GLPQSC-----SYGSDVTSG-----XELSDSCGAGGGRL 1158
Qy 1232 GTPADVLYGTITRIIGEDSPRLDRGREDLSLPKHVIVEGKHVLSYEGGMSVTOCSK 1291
Db 1159 GRAAKHRS-TR-----ANSRQERASRPLTILNVCNTGDK-----WVECOL 1201
Qy 1292 EDGRSSGPPHETAAKRTYDMQEVGRVIRASSASIEGLMGRAIPPERHSPHLLKEQHHI 1351
Db 1202 E-----THNKMVTFKFD-LDGDAPDEIATVMVE-----1229
Qy 1352 RGSITQIPRSVVEAQEDYLREAKLLKRE-----GTPPPPPPSRDL---TEAVKT 1399
Db 1230 HDFILOARETFIEQKQVMDKAEADMLSEDTADRGSDPGTSPHLSCTGLGTGEESKQS 1289
Qy 1400 QALGPL-KLKPAHEG-----LVATVKEAGRSIHEIPREELRHTPELPLAPRPLKEGSIQ 1454
Db 1290 QANAPVYQONVLHTKRWFIICPVAE-----HPAP-EAPESPPPLPLSSLP-PEAS-QDS 1341
Qy 1455 TPLKYDTGASTGS---KKHVRSLIG---SPGRFTPPVHPDLVMDADALERACVYESLK 1509

Db 1342 APYKQDLSSEKQPSFLASQQLLSQAGSNPFGAPPAPLAPSSPPVTALPQDGAAPATSTM 1401
Qy 1510 SRP--GTASSSGSGIARGAPV-IVPELGKPRQSPLTVEDHGAPFAGH-LPRG-SPVTMRE 1564
Db 1402 PEPASGTASQAGP---GTQGLTSEL--ETSQPLA-ETHEAPLAVOPLVVLGAPCT---1452
Qy 1565 PTPRLOEGSISSSKASQDRKLTSTPREIAKSPHSTVPE---HHPHPIPSYEHLLRGVSGV 1621
Db 1453 PAPE---AASTRDASAPREPLPPAP-EPSPHSGTQPQALGQAPLPLP---AAGVAV 1502
Qy 1622 DLYRSHIPLAFDPTSPRGIPLDAAAAYILPRHAPNPTPHLYPPYLIR-----1671
Db 1503 SLATSQLP-----SPPLG-----PTVPP-OPPSALESDEGSPPRV 1537
Qy 1672 GYDPTA-ALEN-RQTIINDYITSQQMHNTATAMAQR---ADMLRGLSPRESSLALNY 1724
Db 1538 GFVDSITKSDEKLTULYQEHVPTSSASAGTPVEVGDRDFTLEPLRGDQR-----1589
Qy 1725 AAGPRGIIDLSQVPHLEFVLPPTPGTPATMDRLAYLPTAPQPFSSRSHSPLSPGGPTH 1784
Db 1590 -----SEVCGDLALPPVPKAVSGRVQL-----POPLVEK---SELAPTRGAV 1630
Qy 1785 LTKPTTSSSERERDRR---ERDREREKSIILTSTTTVEHADIPRG---TEQSSGS 1837
Db 1631 MEQGTSSSMTAESPRSMGLGYDRDGRQVASDSHVVPQDVPFVFPARVEPTDRDGE 1690
Qy 1838 SGSS-----GGGGSSSRSPASHSHAHQHSPISPRTQDALQORPSVLHNTGMKGIT 1888
Db 1691 AGESSABPPSDMTVGQASHPQT-----LGARALGSPKRP-----1728
Qy 1889 AVEPSKETVLIRSTSTSPFVRPAATFPFATHCPLGGTLGVPYPTLMEVPLLPKEA-PRVAR 1947
Db 1729 -----EQQDVSSPAKTGFRFSVVS-----TQDEM--TLASPHLSRYSAPPDVL 1770
Qy 1948 PERPRATGHAFLAKPARSGL-----EPASSPKSGSEPRPLVP-----PVSGHAT--1993
Db 1771 DEAPSSPDVLAVERAQTASSIEVGVEPVSDDGDEGRARPPVQKASLPFVSAGD 1830
Qy 1994 -IARTPAKNLAPHA---SPDPPA-----PPASADPHREKTQSKPFSIQEELRSLGY 2043
Db 1831 FVKATAPLORPSRAGSLGPETPSRVGMKVPTISVTSFHSQSSYISSDNDELE-----1884
Qy 2044 HGSSYSPEGVEPVSPVSPSLTHDKGLPKHLEELDKSHLE--GELRPKQPGFVKLGGEAA 2101
Db 1885 -----DADIKKELQSLREXHLKEISELQSOQKQEI-----EAL 1917
Qy 2102 HLPHLRPLP-----ESQPSPPILQTAPGVYKHQVRVTLAQHISEV 2142
Db 1918 YRRLKGLPLPNVNGFFHTAPTGTRRRKTSKGLKAGLLN--PLVRQLKVASSTGHLAD-1974
Qy 2143 ITODYTRHHPQOLSAPLAPLYSPFGASCPLVDLRRP-----PSDLYLPPPDHGAAPAR 2195
Db 1975 SSRGPPAKDPAQASVGLTA---DSTGLSGKAVQTQPCSVRASLSSDICSGLASDGGAR 2031
Qy 2196 GS-----PHSE--GGKRSPEPNKTSVLGGEGDIEBPVSPPEGMTEPHGHSASVYPLLYR 2247
Db 2032 GQGTVVVHTPSRVTVYKSSSKPRARFLSG-----PVS-----VSIWSALKR 2072
Qy 2248 ---DGEQTESRMCSKSPGNTSQP-PAFFSKLTESNANMVKSKOEINKKLNTNRNEPE 2303
Db 2073 LCLGKHSRSTSSSLAPGPPGPQALHVOAQVNNN---NNKGTFTDDLH-----2121
Qy 2304 YNISQGTETFPNPAITGTGLMTRYSOAVOBSHASTNMGLEAIIRKALMGKYDOWEESPPL 2363
Db 2122 -----KLVDWTSKTVGAQOL--KPTLNQLKQTKQLQDM 2153
Qy 2364 SANAFNPLNASLPAAMPITAADGRSDHTLTSPGGGKAKVSGRPSRKAAPSAGLAS 2423
Db 2154 EAQAG-----WAAPGEA-RAMTAPRAG-----VGWP-----RLPPAPGPLS 2188
Qy 2424 GDRPPSVS---SVHSEGDCHNRRTPLTNRVWEDRPSSAGSTPPY 2464
Db 2189 TTVIPGAAPTLVSPTPGSCGR-----AVSTPSTY 2218

RESULT 55

Q9VH10	PRELIMINARY; PRT; 3111 AA.
AC	Q9VH10;
DT	01-MAY-2000 (T-EMBLrel. 13, Created)
DT	01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE	CG3996 protein.
OS	CG3996.
GN	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7221;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Berkely;
RX	MEDLINE=20196006; PubMed=107311132;
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeon K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C., Ferraz C., Ferreira S., Fleischmann W.,
RA	Feiler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harzick N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA	Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA	Palazolo M., Pittman G.S., Pan S., Folland C., Scheeler F., Shen H.,
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT	"The genome sequence of Drosophila melanogaster.";
RL	Science 287:2185-2195(2000).
DR	EMBL; AE003686; AAP54511.2; --
DR	FlyBase; FBgn0037800; CG3996.
GO	GO: 0005634; C:nucleus; IEA.
DR	GO: GO:0003677; F:DNA binding; IEA.
DR	InterPro; IPR001005; Myb DNA binding.
DR	InterPro; IPR000195; RbGAP_TBC.
DR	PFam; PF00566; TBC; 1.
DR	SMART; SM00164; TBC; 1.
DR	PROSITE; PS00037; MYB_1; 1.
DR	PROSITE; PS50086; TBC_RbGAP; 1.
SQ	SEQUENCE 3111 AA; 349635 MW; EDA9A5FD38115773 CRC64

Query Match 3.5%; Score 458.5; DB 5; Length 3111;
Best Local Similarity 18.1%; Pred. No. 2.5e-13;
Matches 534; Conservative 376; Mismatches 1061; Indels 975;

QY 863 KSECT---EEAEKGAKGDAEAA-----BATAEGAL-----K 892
 Db 1550 LQRTPMGDETSAPANTAYRAANHEWQOBWVRFSDDLPSRADLENRLQFLERQLYKK 1609
 QY 893 AEKKEGSGRATTAKSSGAPQDSSSATCSADEVDEAEGGKNRLLSPRPSL----- 944
 Db 1610 FYKQRCASDSVASKVLPEDQSTSRQARKQ--EAEGOLEQRVLALEKQLSNSLKLL 1667
 QY 945 -----LPTGDPNANASQKPLDLKQLKQRAAAIPIQVTKVHEP-----P 985
 Db 1668 EAMERHRSADDSGSPRLSTETIDATGKGLVRVTONIGELEVEDAHKPINISINIOMV 1727
 QY 986 REDAAPTAPAPPPONLQPESDAPQOPGSSPRGKSRSPAPPADKEAFAAQAQKLPGD 1045
 Db 1728 NKDESQPKGESKFTTDLRLEOLEQOALLEERAKNGS--IPENEVLEERPEKLEEK 1785
 QY 1046 PPCWTSGLFPVPPREVIKASPHAPDSAFSAPPGHPLPLGLHDTARPVLPRPPTSNP 1105
 Db 1786 DSC-----KKQKCNCHQHVKGDEVEKTEIPAD-----RKIEP----- 1818
 QY 1106 PPLISSAKHPSVLE-----ROIGALSQMSVQ-LHVPYSEHAKAPVG 1146
 Db 1819 -----ASAKETKTLENVEKAOAKTRAKVVDTEKSVKQDNNAVTEKSVQDQNVVVDKADRKI- 1873
 QY 1147 PVTWGLPLPMDPKKLAPPFGVKQQLSPRGQAGPPESLGVPTAQAEASVLRGTALGSPGG 1206
 Db 1874 -----LQKOKSPAAGKSEDTKQSGKKESED--IKQASEA----- 1908
 QY 1207 SITKIPTRVPSOAIYRSGITHGTPADVLYKGTITRIIGEDSPRLDRGRDSDLPKG 1266
 Db 1909 -----PRAGASKETS-----TRGKPSE-----TKL---EKPTTKESVLKETFPKK 1945
 QY 1267 HVIYEGKCHVLSYEGGMSVTOCKEDGRSSGPPHETAAPKRTYDMWGRVGRGAISSAS 1326
 Db 1946 ENL-ESEKPKSENEATETQKSETPTVAVSPKESKVKQMTKEK-----TIKDSS 1999
 QY 1327 TEGLMGRAI-----PPERHSPHLKEQHHRIGSITQIGPRSYVEAQBLYL----- 1372
 Db 2000 SKELPEKMNINSTDVGPMDPNGKTVVLLMDNEH-RASKVRLTRANTEEDLEDLQALEKQ 2058
 QY 1373 -REAKLLKREG---TPPPPPSRDLUTEAYKTOALGPL-----KLKPAHEGLVATVKE 1420
 Db 2059 LNDRLNVKSEGRLLRVDPKPSAEQVE--OTQAISDLTKIEDFTSAKPEEE-----NPKE 2112
 QY 1421 AGRSITHEIPRELHRTPELPAPRLKEGSIQTGTPFLKVDTCASITGSKKHDSVRLSISP 1480
 Db 2113 AKEDKPEPEE-----PEDFDGPNV--KHLKKTIVYLP 2146
 QY 1481 G-----RTFPPVPHLDVNMADARALER-----ACYEESLSKSRPGTASSS 1518
 Db 2147 STKELESFRSLERQIKLEEDVEKIDVEQRLNEIERKIKLOVLSHSEKDLNKLYLELCEGK 2206
 QY 1519 GSGIARGAPVIVP---ELGKPR-----QSPLYYEDHGAFAGHLPRGSPVT 1561
 Db 2207 GLDDDEPVVETPTKEAIIITARDSRSPGRKALATKSPYTSRSPKATI--KTPTSPT- 2263
 QY 1562 MEPE-----TPRLEQGSLSKASODKLTSTP-----REIAKSPHSTVPHEHPH--- 1606
 Db 2264 -RKPIIKGPTSPSKSAKSPYTSRNRQSPSTRSPERKSKSPYTSRARKPHND 2322
 QY 1607 -PIS---PYEHLRLGVSGVDLYRSH-----IPLAFDPTSIPIRGIPLDAAAAYLPRHAP 1657
 Db 2323 LPISDDLEYKYRV---LDLVRSKSENKAKRMNDPNRKPAIHPLEMI-----LSP 2369
 QY 1658 NTTYPHLYPPYLRGYPDTAALENQTIINDYITSQOMHNTATAMAQADMLRGLSPRE 1717
 Db 2370 SPDADAI-----PTTGELEHRIRLVDEKLKSP-----AKTRSKSRSP-- 2408
 QY 1718 SSLALNAAAGPGIIDLISOVPHLPVLPPTGTPATAMDRLAYLPTAPOP----- 1767
 Db 2409 -----TIEDIKQKWRDEKKTPTVHNLERIYVSSPCRPPTABELEERI 2453
 QY 1768 -----FSSRHSS-----SPLSPG----- 1780

Db 2454 RILEQEHKFDFTQDYKAFNOKLKDVISFSLSPDEFRAAKREOSPRRHGPTTPKSALR 2513
 QY 1781 -----GPTH---LTKPTTTSSSE-RERDRDRDRDREREKSIILSTTTV----- 1821
 Db 2514 RDDFDGSGYCGTHPTSTLYRTPSKVIRFRDEDEDEDEQFEAPKPSRQISDRMWGTHDV 2573
 QY 1822 -----BHAPIW-----RPGTEQSSGSGSGSGSGS-----SSRPASHSHAHQHP 1862
 Db 2574 LDCLTENTKILQRLKLTADQPSATRSYASSTEGDLGSLRLMRETSPTTRTGTHGVP 2633
 QY 1863 ISPRTOALQORPSVLHNTGKGIITAVEPSK-----ATHCPGLGTLGVTPLMEPVLLKPEAP 1943
 Db 2634 L--RTGENINDRSLSSIKNS-IKSIDTLCB-EKPYQKEKQORYIDSLFTSLHFKSSSL 2689
 QY 1896 -----TVLRSTSTSSPVPRPAATPPP-----PESIQLLELRSLG 2042
 Db 2690 EDLSLSRSLSRSRGRSIRHSGDYAPSRVISEHRSLS-LSADS-----RRSP 2736
 QY 1944 RVARPERPRADTGHAFKAK--PPARSGLPEPASSPKSGSPRPLVPVPSGHATTIARTPAKN 2001
 Db 2737 LGNRDTSPLHHRSHRDISRELSPRRRLEEDDEERKDRES-----SRVRDN 2783
 QY 2002 LAPHASPDPPAPPASAS-----DPHREKTQSK----- 2842
 Db 2784 LLENYFADNRSELSSGSSSLTGFNHVKDQLEETCAKADDDRRSACRTPLS-HPYESRTTA 2842
 QY 2043 YHGSSYSPEGVEPVSPVSSPSLTHD--KGLPKHLEELDLSHLEGLRKPQPGVPLGGEA 2100
 Db 2843 TRISHDTPVOI-PTNPAGSATAIDSFPSPVSPQVPDPVH-----RSPGGA--GGTP 2892
 QY 2101 AHLPL-----HLRPLPESQSSSPLQLQAPGVKGHQRV 2132
 Db 2893 LYQPKLEIRHTTWTSTFYDRFLTEKQIERQTHSRP-----PSRSPVSPVSPAKSYVEL 2947
 QY 2133 VTLAQHISEVITODYTHHPQQLSAPLAPLYSPGASCPLVLDLRPPSDLYLPPDPHGA 2192
 Db 2948 CSTSGTSTTATSTSTSSF--MSSSYAGSFSLPSAS-----NFSYLNLP----- 2989
 QY 2193 PARSGPHSEGGKSPENKTSVLGGGEGDIEPVSPPEGTEPGHRSRSAVYPLLYRDGEOT 2252
 Db 2990 --GS-----GSGSGISSISP----- 3002
 QY 2253 EPRMGSKSPGNTSQPPAPFSLKLTENSAMVSKKQKQINKKLNTHNREYVILSQPTE 2312
 Db 3003 ---RASCSDLRSITSGTSTSTSTSTSYV-----PYNFTSSFTS 3040
 QY 2313 IFNMPAITGTLMYRSQAEHASTNMGLEAIIRKALMGKYDQWESPLSANAFNPLN 2372
 Db 3041 RLNDPIITSTTSVSTSSLTSTHSTGTVNPMMSFTLRE-----PLASSSLGGSS 3087
 QY 2373 ASASLP 2378
 Db 3088 ASPLLP 3093
 RESULT 56
 QP2R6 PRELIMINARY; PRT; 1566 AA.
 AC QP2R6
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE RERE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20275632; PubMed=10814707;
 RA Yanagisawa H., Bundo M., Miyashita T., Okamura-Oho Y., Tadokoro K.,
 RA Tokunaga K., Yamada M.,

RT "protein binding of a DRPLA family through arginine-glutamic acid
 RL dipeptide repeats is enhanced by extended polyglutamine."
 Hum. Mol. Genet. 9:1433-1442(2000).
 CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AB036737; BAA95898.1; -.
 DR Genew; HGNC:9965; RERE.
 DR GO; GO:0005634; C:nucleus; NAS.
 DR GO; GO:0005515; F:protein binding; NAS.
 DR InterPro; IPR002951; Atrophin.
 DR InterPro; IPR001025; BAH.
 DR InterPro; IPR000949; ELM2.
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR InterPro; IPR000679; Znf_GATA.
 DR Pfam; PF01426; BAH; 1.
 DR Pfam; PF01448; ELM2; 1.
 DR Pfam; PF00320; GATA; 1.
 DR Pfam; PF00249; myb_DNA-binding; 1.
 DR SMART; SM00439; BAH; 1.
 DR SMART; SM00717; SANT; 1.
 DR SMART; SM00401; Znf_GATA; 1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 1566 AA; 172336 MW; 4047F1530F198E2C CRC64;

Query Match 3.4%; Score 452.5; DB 4; Length 1566;
 Best Local Similarity 19.0%; Pred. No. 2e-13;
 Matches 403; Conservative 205; Mismatches 660; Indels 853; Gaps 90;

QY 204 AAKPEPEKVPSPPIESK-----HRSLVQIYIDENRKAEEAH 242
 DB 158 ASQPPQHLSEAGRGVSGKRDHLLMNVKYYRQSEVPDSVYQHLVQDRHNDSGRELV- 216
 QY 243 RILEGLQVQLPLYNQPSDTRQYHENIKINQAMKLLILYFKRNHARKQWKOK----- 297
 DB 217 -ITDPVKNREL-FISDVYDT--YHA-----AALGKCNISHFSDFIFAAREFKARVDSFF 267
 QY 298 FCQRYDQLEALEKKVERIENPRRRAKESKVREYYEKQFPEIRKQRELOERMOSRVQOR 357
 DB 268 YILGYNPETRLNSTQGEIRVGPSHQAKLPDLQPPSPDGTVTQHEEL-----VWMP 320
 QY 358 GSG-----LSMSARSEHEVSEIDGLSEQENL-----EKQMLQAVIPWMLYDADQOR 406
 DB 321 GVNDCDLLMYLRAASMAAFAGMCDGGTDCGVAASRDDTTNLNALTLSHESGYDAGKA- 379
 QY 407 IKFINMGLMADPMKVYKROVMNWSQEKTFREKFMQHPKNGFLI-ASFLEKKTVAE 465
 DB 380 -----LQRLVKPV-----PKLIEKCTEDEVRKFGVGLRQYQGNPFIRKELLPNKETGE 430
 QY 466 CVLYYYLTGKN-ENYKSLVRSYRRRGKSQQQQQQQQQQQQQQQQQQQQPPRSSQBEKDEKE 524
 DB 431 LITFYVYWKTPEAASSRAHRRRQAVFR-----RQI 462
 QY 525 KEKEAKKEEKEPEVENDKE--DLLKBTDDTSGEDNDEKEAVASKRKTANSQQRKRGRI 582
 DB 463 KTRTASTVNTPSRPPSEFLDLSASRDDFDSEDSQE-----LKGYA 506
 QY 593 TRSMANEANSBEAITPQOSAEALSAMELNESSRWTEEMETAKGLLHGRNWSAIAARMVG 642
 DB 507 CRHCFTTTSKD-----W-----HHG-----G 522
 QY 643 SKTVSQCKNFYFNVKRQNLDEILQHKLMKEKERNARRKKKAPAAASEEAPPPVVED 702
 DB 523 RENILLCTDCRIHFVKYGELEPPI-----EKPVDP-----PPFMKPKVKEE 562
 QY 703 EEMASGVSGNEEMVEBAEALHAGSNVPRGCGSPATVNNSDTESIPSPHTEAAKDT 762
 DB 563 DD-----GLSGKHSMRTSR-----RGSMSLTRSGRKKQKQSPDCRT-----SPINEDIRSS 609
 QY 763 GONGKPKPATLGADGPPPGPTPPRRTSRATIEPTPASEATCAPTPPAPSPSAPPVV 822
 DB 610 GRNSPSAASTSSND-----SKA-----ETV 629

QY 823 PKESKBETAAAPVVEGEQKPPAAELAVDTGKAEEPVKSECTEAB-----EGP 874
 DB 630 KKSARKVKEEASPLKSNKQR-----EKVASDTEADR-TSSKKTQTEISRPNSPSEG- 683
 QY 875 AKGDAEAAEATAEGALKAEKKEGSGRATTAKSSGAPQDSDSSATCSADEVDEAGDCK 934
 DB 684 -EGSSDSRSVNDGSS--SDPKDIDQNRSTSPSPSQDNESDSDSAQQ----- 731
 QY 935 NRLLSPRPSLLTPTGDPRA--NASPQKP-LDLKQLKORAAAIPTQVTKVHEPPREDAA 991
 DB 732 QMLQAPQALQATGVTAPSSAPPGTQLPPTGPTSATAVPP-----QGSPTASQAP 785
 QY 992 TKP-APPAPPPQNLQESDAPQPGSSPRGKSRSPAPPADKFAFAEAQKLPGDPPCWT 1050
 DB 786 NQQAAPTAPVPHTHIQ-----QAPALHPQ-RPPSPHPPH----- 819
 QY 1051 SGLPPVPVPREVIKASPHADPSAFSAPPQGHPL-----PLGLHD-TARPVL--PRPPTIS 1103
 DB 820 ---FSPHPFLQPLTGS--AGQPSAPSHAQ--PLHGQGPFGPHSLQAGPLLOHPGFPQPF 872
 QY 1104 NPPLIISAKHPSVLERQIGAISQGMVQLHVPYSEHAKAPVGPVTMGLPLPMDPKKLAP 1163
 DB 873 GLPQASQGGAP-----LGTSPAAAYPHT-----SLQLPASQAL-- 907
 QY 1164 FSGVQEQQLSPRQAGPPESLGVPTAQEASVLRGTALGSGVPGSITKIGIPSTRVPSDSAI 1223
 DB 908 ----SQQPPREQLPFGPLMPH-----IKPPTTPIPOLAA- 941
 QY 1224 TYRGSITHTGVPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEGKKGHVLSYEGG 1283
 DB 942 -----POAH-----KHPPHLSGSPS 956
 QY 1284 MSVTCQSKEDGRSSGPPHETAAPKRTYDMMEGRVRAISSAIEGLMGRAIPERHSPH 1343
 DB 957 FSM-----NANLPPPPALKP-----LSSLSTHH-----PSAHPPP 987
 QY 1344 -HLKEQHIRGSIQTGIPRSYVEAQEDYLREAKLKRGGTPPPPPSRDTEAYKTOAL 1402
 DB 988 LQLMPQ-----SQPLPSS--PAQPPGLTQSNL-----PPPPASHPT-- 1023
 QY 1403 GPLKKAHBEGLVATVKEAGRSIHEIPREELRHTPELPAPPLKEGSITOGTPLKYDTG 1462
 DB 1024 -----GLHQV-----APQPPFAQHFPVPGPPPTTPT-- 1051
 QY 1463 ASTTGSKKHDVRSLLIGSPGRTPPVHPLDMVADARALACRYESLSKSRPGTASSGSGI 1522
 DB 1052 -----CPSTSTPPAGP-----GTSAQPPC-----SGAASGSGI 1080
 QY 1523 ARGAPVIVPELGKPRQSPPLYEDHGAFFAGHLPRGSPVTWREPTPRLOEGLSSSKASQD 1582
 DB 1081 AGGS-----SCPL-----PTVQIKEALDDAEPE- 1105
 QY 1583 RKLSTPRETAKSPHSTV---PEHHPIPISPYEHLLRGVSG---VDLYRSHIPLAFDPTS 1636
 DB 1106 ---SPPPPPRSPSPSEPTVTDTPSHASQASARFYKHLDRGYNASCARTDLY--FMPLA- 1155
 QY 1637 IPRGPLDAAAAYVLRHLAPNPTYPHLYPPYLIRGVPDAALENRQTIINDYITTSQOMH 1696
 DB 1156 ----- 1155
 QY 1697 HNTATAMAQRADMLRGLSPRESSIALNYAAGPRGIIIDLQVPHLPVLVPPPTGTPATMD 1756
 DB 1156 ---GSKLAKKEEAIEKAREAB----- 1175
 QY 1757 RLAVLPTAPOFFSRHSSSPLSPGPGPHLTKPTTTSSERERDRDRDRDREREKSILT 1816
 DB 1176 -----OKAREEREKEKEKERERERE----- 1199
 QY 1817 STTTVEHAPIWRPCTEQSGSSGSGSGGSSSRPASHAHQHSPISPRTQDALQORPS 1876
 DB 1200 -----REARAKAASSAHEGRSLDFQLSGFGHMRRSPSEPPPTTAAVPPY 1245
 QY 1877 VLHNTGMKGIITAVEPSKPTVLRSTSTSSPVRPAATPPPAPATHCPLGGT-----LDGVY 1929

```
Db 1246 IGPDT--PALRTLSEYARPHVMSPTNRNHPF-----YMLNPTDPLLAHYHMGLY 1293
Qy 1930 ---PTLMEPVLLPKF-APRVARPERPRADTGHAFLAKPABSGLEPASSPSKSGSEPRPLV 1985
Db 1294 NVDPTIRERELREIREIREIRERELRMRKPGFEVKKPELDPJLHPAAN-----1342
Qy 1986 PPVSCHATIARTPAKNLAPHASPDPPAPPASADPHREKTQSKPFS-----IQEELRS 2040
Db 1343 -----PMEHFAHSAITIPP-----TAGPH-----PFASHPGLNPLERER 1378
Qy 2041 LGYHGSSYSPE-----GVPEFVSPVSPSTHSD-----KGLPKHLEBDSKHLEGE 2085
Db 1379 LALAGPQLRPMSYDRLAERIHAERIASLTSDDLARLQMFNVTPHHQH---SHISH 1435
Qy 2086 LRPKQPGVKLGGEAAHLPHURPLPESQSPSPILLQTAPGVYKQHVVTLLAQ--HISEVI 2143
Db 1436 LHLHQDPLHQGSAGPVHPLVDPL-----TAGPHLARFPYPGTLPLNLLGQPPHEHEIV 1490
Qy 2144 TQDTRHHQQLSAPLAPLAPLYSFCASCVPVLDLRPPSDLYLPPPDHCAP-----2193
Db 1491 -----RHP-VGTYPRDL---PGA-----IPPMASNAHQLOAMHAQSA 1525
Qy 2194 -----ARGSPHSEGG 2203
Db 1526 ELQRLAMEQQWLHGHPHMG 1546

RESULT 57
QYR40
ID Q7YR40 PRELIMINARY; PRT; 2171 AA.
AC Q7YR40;
DT 01-OCT-2003 (TreeBLrel. 25, Created)
DT 01-OCT-2003 (TreeBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TreeBLrel. 25, Last annotation update)
DE Homologue to Drosophila photoreceptor protein calphoton.
GN KIA0170.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22709134; PubMed=1279463;
RA Anzai T., Shiina T., Kimura N., Yanagiya K., Kohata S., Shigenari A.,
RA Yamagata T., Kuleki J.K., Naruse T.K., Fujimori Y., Fukuzumi Y.,
RA Yamazaki M., Tashiro H., Iwamoto C., Umenaha Y., Imanishi T.,
RA Meyer A., Ikeo K., Gojobori T., Bahram S., Inoko H.;
RT "Comparative sequencing of human and chimpanzee MHC class I regions
RT unveils insertions/deletions as the major path to genomic
RT divergence."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7708-7713(2003).
DR EMBL; AB100084; BAC78176.1; -.
KW Receptor.
SQ SEQUENCE 2171 AA; 235386 MW; 8D8F30C9CDEF7114 CRC64;
```

```
Query Match 3.4%; Score 452.5; DB 6; Length 2171;
Best Local Similarity 19.8%; Pred. No. 3.1e-13;
Matches 495; Conservative 290; Mismatches 897; Indels 817; Gaps 116;

Qy 18 RYPHSLSYFV-QIARTHTDVGLEYQHS--RDYASHLSFGSIITQQRRLPSLLSEFQP 74
Db 58 RMPDCSVALPFSISKQAEIILAWDKAPILRDCGS-LNGTQILRP---PKVLSPGVS 112
Qy 75 GNERQEL-----HLRPESHY---LPELKSMEFIESKRPLELLPDLRPSLLATG 127
Db 113 HRLRQELTLFADLLLCQHLRDLVSFLFVSRLPTVEETPRVGGGTPQRL-----LLAED 167
Qy 128 QPAGSEDLTKDRSLATGKLEFVSPSPPTDDELE-----LVPPRL-----SKEELIQNWD 177
Db 168 SEEVDFLSERRVWKSRRTSSSVIPSDSEGHSPVLGGLGPPFAPNLNSDITDVEEQQ 227
Qy 178 RVDREITWVEQISLKKKQKQLEEEAAKPPPEKPVSPPTTESKHSRLVQIIVDENRKK 237
```

```
Db 228 PATETASSARRGATVEAKQSEAEVVTEIQLEKQDP-----LVKERDDDTKVK 275
Qy 238 AEAARILEGLGQVELPLYNQ-----SDTROXHENIKINQAMRKKLILYFKERNHARKQ 293
Db 276 RGAEN-----GVVPAGVILERSQPPGEDSDT-----301
Qy 294 WKOKFCORYDQMLEALEKKVERIENNRARRAKESKRYEYKQFPPIRKORELOER-----349
Db 302 -----DVDDDSRPGRPAEVLHRAQAPFGFINSDDAEERIPA 340
Qy 350 -----MQSR-----VGQSG-----LSMSARSEHEVSEIIDLGLSEQENLEKQRL 392
Db 341 TPVVIWMKKEKIFHGVTGPGAPGLAHLQESQAGSDTVEE---GKAPQ-----387
Qy 393 AVIPPMLYDADQRIKFINMGLMADPMKYKDRQVNMWMSOEKETFREKFMHOKNFG 452
Db 388 -----AVPLEKSQASWINSDTDEEVEVSAALTLAHLK---420
Qy 453 LIASFLERKTVABCVLYLYLTKKENYKSLVRRSYRRRGKSQQQQQQQQQQQQQQPM 512
Db 421 -----ESQPAIMNRDAEDM 435
Qy 513 PRS---SQEEKDEKEKEAEKEEKEPEVENDKEDLLK-----EKTDDTSGE 556
Db 436 PORVLLQRSQTTTERSDTDVEEELPVEN-REAVLKHQHTKTRALVRAHSEKQDPPFGD 494
Qy 557 DNDEKEA-VASKGRKTANSQGRKGRITRSMANEANSEAI-----TPQOS--601
Db 495 SDDSVEADKSPGIHLERSQASTTVDINTQVEKEVPPGSAIHHKKHQSVEGTNQTDVK 554
Qy 602 -----AELASMELNESRW-----TEEMETA-----KKGLL-----BHRNWSA 636
Db 555 AVGSPAKLLVVSLEEA--WPLHGDCETDAEDTSLAASAVADVRSKSLPAEGDAGAEWAA 612
Qy 637 I-----ARMVSK---TVSQCKNFYNYKKRONLDEILQOHLKMKERNARKKKKAP 687
Db 613 AVLKQEAHEVGAQGGPPVAQVEQDL--PISRENLTDLVVDTDTLGESTQPR-----663
Qy 688 AAASEEAAFPVVEDEMEASGVS-----GNSEEMVEEA-----EALHASGNEV 731
Db 664 -----EGAQVPTGREHQHVGKTDSEDTYGDSEDLDLQATCFLENQGLEAVQSMEDP 718
Qy 732 PR-----GECS-----GPATVNSSDTESIP-SPHTEAAKDT 762
Db 719 TOAFMLTPPQELGPHSCSFOTTGLDBPWEVLATQPPCLRESEDSQTPTFDTHLEA---774
Qy 763 QONGP--KPTATLQADGPPPGPTTPPRTSRAPTEPTPAS-EATGAPTPPAPSPSAPP 819
Db 775 --YGFCLSPRAIPGDOHP-----ESPVHTEPMGIQGRGRTVDKVMG-----815
Qy 820 PVVPEKEEETAAAPVVEEGEOKPPAAEBELAVDTGKABEPVKSECTEEAEFGPAKGD 879
Db 816 --IPKETAARVGPGRGLERETEKLLP---ERQDVTGEBELTKGQDRQKOLLARDTQ 870
Qy 880 AEAATAEAGALKAEKKEGGSGRATTAKSGAPQSDSSATCSADVEDAEAGGDKNLLS 939
Db 871 QRESDKNGESASPERDRESLKVEIETSEEIQEKQVQKQTLPSKAFE-REVERPVANRECD 929
Qy 940 PR-----PSLL-----TPTGDPANASPOKPLDLKQLKQRAAAATPPIQVTKVHEPPREDA 989
Db 930 PAELEEKVPKILBERTQGEPEGSGDQK-----GQASSPIP-----BFGVE--972
Qy 990 APTKAPAPPQPNQLOPES-DAPOQSGSPRGKSRSPAPPADKFAFAEAQKLP-GDPP 1047
Db 973 AGDLPGTSAVPTSGSGSGGSGSPVSRHHQGLLNCKMPPAEKASRIRAEKVSRRGQDE 1032
Qy 1048 CWTSGLPFPVPPREVIKASPHADPFSAFSVAPPGLPLGLHDTARVPLPPPTISNPPP 1107
Db 1033 SPDACLPPTV-----PEAP-----APPOKP-----LNSQSKHLAPPPLLS---P 1069
Qy 1108 LISSAKHPSVLE-ROI GAISQMSVOLHVPVSEHAKAPGVPTWGLPLPMDPKKLAPPSG 1166
```

Db 1070 LLPSTK-PTVRKTRDGS-----QEAPEAPLS-----SELEPH- 1102
Qy 1167 VKQEQLSPRGQAPGESLGVPTAQBAQSVLGRGTALGSGVSGSITKIGPSTRVPSDAI--- 1223
Db 1103 -----PKPKIRTKSRMTTFPATSAPEHPSTSTAQVPTPKPTSQATRSR 1149
Qy 1224 TVRSGI---THGTP-ADVLYKGTIT-RIIGEDPSRLDRGRDLSLPKGHVIYEGKKHVL 1278
Db 1150 TNRSSVKTPSEQVPTAPELQPCSTSDQPVTSFTSQVTRGR----- 1191
Qy 1279 SYEGCMSVTCQSKEDGRSSGPPHETAAPKRYDMMGRGVGRAISSASIEGLMGRAIPE 1338
Db 1192 -----SRSSVKTP-ETVPTALELQFSTSTDRPVT-----BPT 1224
Qy 1339 RHSPHLLKEQHIRGSIITQIGIPRSVYEAQEDYLRREKLLKREGTPPPPPSRDLTEAYK 1398
Db 1225 SHATGRKN-----RSSVKTP-----PVVPTAPELOPSTS 1255
Qy 1399 TQALGPLKPKAHEGLVATVKEAGRSIHIEIPRELRHTPEL-----PIAPRPLEKGS 1450
Db 1256 TDQ--PVTSEPTYQ---ATRGKRNSSVKTPPEPVVPTAPELOPSTSTDPVTPKPT--- 1306
Qy 1451 ITQGTPLKYDTCASTGSKKHVDRSLIGSPGRTTP-----PVHPLDVMADAR 1497
Db 1307 -----SRTTRKTN--MSVKNPSTVPIAPELPSTSTEQVTP----- 1344
Qy 1498 ALERACYESLSRPGTASSGG-----SIARGPVIVPELGPQSPPLYEDHGAPPAGH 1553
Db 1345 -----PFTSATRGKRNSSGKTPETILVPTAPKLEPSTS--TQVPTPEPTSQATGR 1395
Qy 1554 LPRGSPVTWREPTPLQBGSLSSSKASQDRKLTSTP-----REIAKSPHSTV--- 1600
Db 1396 TNRSSVKTPETVPTAPELQST---STDQAVTPKPTSRTRTNRMSVKNPSTVPIA 1452
Qy 1601 PEHHP-----HPISEYHLLRGVGVLYRSHIPLAFDP-TSIPRGILPDAAYLYLRH 1654
Db 1453 PELPSTSTEQVTP-EPTSRATRG-----RKNRSSGKTPETILVPTAPKLEPSTSTDP- 1505
Qy 1655 LAPNPTYPLHYPIYLRGVPDPAALENRQTIINDYITSOQHNTATAMAQRADMLRGLS 1714
Db 1506 VTPEPT-----SQATRGTRNRSSVKTPETVTP---TAPELOPSTSD-----QVPT 1548
Qy 1715 PRESSALNAYAGPRGIIDLQVPHLVLPVPTPGTATAMDRLAYLFTAPQPFSSRHS 1774
Db 1549 PEPTSQV-----TRGTRDRSSVK-----TPETV-----VPTAPELOQASSTD 1585
Qy 1775 SPLSPGGTHLTKPTTTSSSERDRDRDREREKSLTSTTTVEHAPIWRP--GTE 1832
Db 1586 QPVT-----SEPTSRIT-----RGRKNRSSVKTPETVPTAPELOPSTSTD 1626
Qy 1833 QSSGSSGSGGGSSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKGIITAVEP 1892
Db 1627 QVTPETSOATRGTRNR-SSVKTPESIVPTAPELQ-----PSTSRNQ-----LVTPEPT 1675
Qy 1893 SKPTVLR---STSTSSVPRPAATPPATHCPGLGTLGCVPTLMEPVLPLKPEAPRVARP 1948
Db 1676 SRATCRTRNRSSVKTPPEVPTAPEPHPT-----TSTDQVPTPKLTSRATR 1722
Qy 1949 ERPRADTGHAFIAPK--PARSGLEPASPSPKSGSEPRPLVPVPSGHATTIARTPAKLAPHH 2006
Db 1723 KTRNSVK---TPKPEVPAASDLEPFTTDOQSVTPEIAQ--GGOSKTLRSTVRAMPVP 1777
Qy 2007 ASPDPAPPASADPHREKTQSKPFSIQEELRSLGYHSSVSPGVEPVSPVSPSLTH 2066
Db 1778 TTYPEQSP-----VTTDQPI-----SPB-----PITQPS-- 1802
Qy 2067 DKGLPKHLELDKSHLEGELRPKPGPVKLGGAAHPLHLPL--PESQFSSPSLLQTAPG 2125
Db 1803 -----IKRQRAAG-----NPGSLAIPIDHKPCSAFLEPKSQASRN---QRNGA 1842
Qy 2126 VKGHQVVTLAQHSITSEVITQDYTRHHPOOLSAPLAPLYSPFGASCPLVDLRPPSDLYL 2185
Db 1843 VRAESLTAIPEPASPOLLE--TPIHASQIQKVEPAGRSRFTPELOPKASQSRKSLATM 1900

Qy 2186 PPPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEGDIEFVPPPEGWTBPGHSRSAYVP-- 2243
Db 1901 DSP-----PHQKQQRGEVSKTVIKEEEE-----DTAEKPGKEEDVVTPKP 1943
Qy 2244 -LLVRDGEQOTPSRMGSKSPGNTSQPPAFFSKLTESNSA 2281
Db 1944 GKRRDQAEENPNIPIRSRLRT-----KLQOESTA 1974
RESULT 58
ID Q9Y5L9 PRELIMINARY; PRT; 2971 AA.
AC Q9Y5L9;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Transcriptional activator SRCAP.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9278407; PubMed=10347196;
RA Johnston H., Kneer J., Chackalaparampil I., Yaciuk P., Chrivia J.;
RT "Identification of a novel SNF2/SWI2 protein family member, SRCAP,
RT which interacts with CREB-binding protein.";
RL J. Biol. Chem. 274:16370-16376 (1999).
DR EMBL; AF143946; AAD39760.1; -.
DR TRANSFAC; T04151; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008086; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF02178; AT hook; 3.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR PRINTS; PR00929; ATHOOK.
DR SMART; SM00384; AT_hook; 3.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 2971 AA; 315639 MW; F9F7EE70304B78A3 CRC64;

Query Match 3.4%; Score 452; DB 4; Length 2971;
Best Local Similarity 19.0%; Pred. No. 4.8e-13;
Matches 587; Conservative 296; Mismatches 1107; Indels 1092; Gaps 130;
Qy 71 EFQGNERSQS---LHLRPESHVLPGLKSEMEFI--ESKRPRLELLPD--PLLRSP 122
Db 91 DFQOEDEEEDDEETIEVEEQEGNDAAQRRLELLRREGELPLELLRSLPQLLEGP 150
Qy 123 LLATQCPAGSBDLTKDRSLATGKLEPVSPSPPTDPELELVPPR---LSKEELIQMDRV 179
Db 151 SSPQTPSSHSDTRDGPGEAAE-----EPQV--LEIKPPPSAVTQRNKPWHPDED 202
Qy 180 DREITWVEQIISKULK---KKQQQLLEEAAPPEPKVPSPPIESKHSLSVLIY---- 231
Db 203 DEEFTANEDEEEDDTTAAEEQLEGEVDHAMELSAREGEL-----SMEELQOYAGA 257
Qy 232 -----DENRKKAAAHRIILGLGP-OVELPLYNQPSDTQVHENIKINQAMRKXLI 281
Db 258 YAPGSGSSEDEDEVDANSDCPEGEVEAEP---PQEDSSSSQSDSVE----- 304
Qy 282 LYFKRRNHARKQKQKFCQRYDQLMEALEKKKVERIENNPRRAKESKVREYEQFPEIR 341


```
QY 1985 VPPVSGHATIAITPAKNIAPHASP---DPPAPPASAS-----DPHREKTSQKPSI 2033
Db 2246 CTPPPAHTP---PQAOTCLVTPSSPLLLGPPSPVPSASVTNPLGLRPEAEALCAQALASP 2302
QY 2034 QELRLSLGYHGSS-----YSPGVPEVS-----PV-----SSPSTHDKG-LPKHLE 2075
Db 2303 ESLELASVASSETSLSLVPPKDLPLVAVELIPVSEKNLSLTPSAPSLTLEAGSLPNGOE 2362
QY 2076 ELDKSHLGEALRPKPGVPLKGGEAAHLP-----HLRPLPESQSSSPLLTQAPGVKXH 2129
Db 2363 QEAPDSASGTTITVLP-----EGEELPLCVSESNGLLEPPSAASDEPLQPLEADR-- 2413
QY 2130 QRVTTLAQHISEVITQDTRHHPOOL-SAPILAPLYSPFGASCPVLDLRRPPSDLYLPPP 2188
Db 2414 ---TSBELTEAKTPTSPKPKQLVTAEEAAPSTSSSATSSP----- 2452
QY 2189 DHGAPARGSPHSEGGKRPENKNTSVLGGEDGIEPVSPPEGMTPEGHSRANVYPLLRYD 2248
Db 2453 -----EGPSPARPPR-----RRT 2465
QY 2249 GEOTEPSRMGSKSPGNTSQPPA--PFSKLTSNSAMVKSKEOEINKLNTNHRNEPEYNI 2306
Db 2466 SADVEIRCOQTCRFG--QPQPKVLRLKPLGRLVTIVVEKELVQRR----- 2509
QY 2307 SQPCTEIFNMPAITGTGLMTRYSAQVQBHASTNMGLEAIRKALMGKYDOWEESPLLS-- 2364
Db 2510 QORGAASTLVFGVSET-----SASPGSPSV--RSMGP-----ESSPPIGGP 2549
QY 2365 -----ANAFNPLNASASLPAAMP--ITAADGRSDHWT 2395
Db 2550 CEAPSSSLTTPPOQFFIARRHIELGVGGSPENGDGALLAITPAVRRRRGPPKKQR 2609
QY 2396 SPGGGKAKVGRSPSRKAKS---PAPG-----LASGDRPPSVSVHSEG 2437
Db 2610 SPADAGVGDVAPSTLKGKNGADVPEPETLIVADVPLPQLIPGQPLGPQPVH--- 2666
QY 2438 DCNRRPTLTNRVWE---DRPSAGSTPPFNPPLMRLQAGVNAS-----PPPGGLP 2485
Db 2667 ---RPNLLSPVKEHRRGPPKARDLP-----GTISSAGDGNSSRTQPPHP 2713
QY 2486 AGSGPLAGPHAWDEEPKLLC 2507
Db 2714 SPLTPL-----PPLLVC 2725

RESULT 59
QY3S1 PRELIMINARY; PRT; 2296 AA.
AC QY3S1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative mitogen-activated protein kinase kinase.
GN PRKNK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon crypt epithelium;
RX MEDLINE=21455683; PubMed=11571656;
RA Verissimo F., Jordan P.;
RT "WNK kinases, a novel protein kinase subfamily in multi-cellular
RL organisms."
RN Oncogene 20:5562-5567(2001).
RP [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon crypt epithelium;
RX Jordan P.;
RA Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ242724; CAB44308.4; -
```

```
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot kinase
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Kinase.
SQ SEQUENCE 2296 AA; 242603 MW; D738FC6C316CA639 CRC64;

Query Match 3.4%; Score 450.5; DB 4; Length 2296;
Best Local Similarity 19.7%; Pred. No. 4.1e-13;
Matches 565; Conservative 257; Mismatches 972; Indels 1069; Gaps 140;

QY 55 PGSIHQPOR-RRPSLLGE-----FQGNERSQELHLRPEHSHSYLPGLCKSEWFEIESKRP 108
Db 11 PGLMEFGRGAGPAGMAEPRAKARPGQRF-----LRSVSVESQDEPFP 55

QY 109 RLELL-----PDPLLR-----PSPLLATQGP-----AG 131
Db 56 GLEAAEAPGPQPQLQRRVLLCKTRLIAERARGRPAAPAAALVAQPGAPGADAG 115

QY 132 SEDITKORSLTGKLEPVSPSPPHTDPELELVPRLSKEELIQNMDRVDRITWVEQOIS 191
Db 116 PEPV-----GTQEP--GPDIAAAVETAPADPGPREAAATVRKEDEGAEEAKPFG 166

QY 192 KLKKKQKQLEEE-----AAKPPPEKPVSPPTLES 221
Db 167 RTRDEPEDEEEDDDLKAVATSLDGRPLKFDIELGRGSFTVYKGLDTWVEVAVWCEL 226

QY 222 KHRSLVQIYIDENKKAEEAHRILEGLCPQVELPLYNQPSDTROYHENIKINQAMKKLI 281
Db 227 QDRKLTGL--ERQFKSEA--EMLKGL-----QHPNIVRFY--DFWESSAKGKRCI 271

QY 282 L-----YFKRNRHARKQWKQKFCORYDQLMEAL-----EKKYVER 315
Db 272 VLVTELMTSGTLTKYLRKFKVMPKVLRSWCR---QILKGLLFLHTRTPPIIHRDLKCDN 328

QY 316 I-ENNRPRRAKESK-----REYVEKOF-----PEIRKQRELOERMQSVGQSGLSM-S 364
Db 329 IFITGPGSVKIGDLGLATLKRASFASVIGTPEFMAPMEYEEHYDESVDVYAFGCMWLE 388

QY 365 AARSEHEVSEIIDGLSEQENLEKQMLAV-IPPMLYD-----ADQ 404
Db 389 MATSEYPSYSEC-----QNAAQIYRVKTCGIKPASEKVDHPEIKELIGECICKNKEER 441

QY 405 QRIKFNNMGLMADPMKY-----YKDRQVNMNMWSEQEKETFFREKFWQHPKQFGLI 454
Db 442 YEIKDOLLSHAPFAEDTGVRVELABEDHGKSTIALRLWVEDPK-----KLKGKPKDNGAI 496

QY 455 A-SF-LERKT---VAECVL---YYLTKQENYKSLVRSYRRR-----GKSO 494
Db 497 EFTFDELEKETPDEVAQEMIESGFFH-----ESDVKIYAKSIRDRVALIQWRRIWPAQ 551

QY 495 QQQQQQQQQQQQQQQQQPMP-----RSSQEEKDEKEKEKEKEKEKEKEKEKEKEKEKE 545
Db 552 PKEQDVGSPDKARGPPVPLQVQVYHAQACQPPPEPE-EPEADQHLHLLPTLPTSATSL 610

QY 546 LKERTDPTSGDNDDEKAVASKGRKTANSQGRKRGRITRSMANEAN-----SEE 594
Db 611 ASDSTFD-SGQGS-----TVYSDSQSSQQSVMLGSLADAAPSQAQCVCSPPVSEG 659

QY 595 AITPQQSAELASMEINLESSRWTEEMETAKGLLEHGHNWSAIAIRMVGSKTVSQCKNPF 654
Db 660 PVLPSQLPSLGA-----QOPTAAPGL-----PVGSVPAPAC----- 691
```

Qy	655	NYKKRQNLDELLOHKLKMEKERNARRKKKKAPAAASEEAAFPVPVVEDEWEASGVSGNE	714
Db	692	 -----PFSLOOH-----FP-----DPAMSFAPV----	709
Qy	715	EEMVEEAEALHASNGNEVPRGECSPATVNNSSDTEIPSPHTEAAKDTQNGKPKPATLG	774
Db	710	 -----LPPFST-----PMTGFGQFAPP-----G	728
Qy	775	ADGPPPGPTP-PRRTSRAPTEP-----TPASEATGAP-----	806
Db	729	 QOPPELAQPTLPQVLAQPFVVPLOQVPPHLPVYLAQFASQ-VGAPAQKLQPLQMPQAPLQF	787
Qy	807	---TPPAPAPSPAP-----PPVVPKEEKESETAAAPVPERGE-----	841
Db	788	LAQVPPQMPPTPVVPPTITPLAGIDGLFPALP-----DLPTATVPVPVPPOYFSPAVILPSL	843
Qy	842	EQKPPRAEELAVDTGKAEE-----PVKSECTEAEEGPAKGDAAEATAEGALKAEKK	896
Db	844	 AAPLPPASPALPQAVKLPHPHPGAPLAMPCKRTIVPNAP-----	891
Qy	897	EGSGRATTAKSSGAPQDSSSATSCADEVDEAGGDKNRLSPRPSLLTPTGDP-----	951
Db	882	-----ATIPLLAVAPGVAALSIHSA--VAQLPGQFVYPAAFPQ---MAPTDVPPSPHH	930
Qy	952	---RANASPKPLDKOLKORAAAIPIQVTKVHEPPREDAAATPKPAPPAPPQNLOPE	1008
Db	931	TVQNRATPPQF-----ALPP-----QPTLPPQPVLPQPTLPPQ	965
Qy	1009	SDAPQCGSSPRGKSRSPAPPADKAEAAEAQKLPDGPWCWTSGLPFVPVPRVVIKASPH	1068
Db	966	FVLPPQF-----TRPPQFVLPQPMPLPQFVLPQF-----ALPVRAPELP---OPH	1008
Qy	1069	APDPSAFSYAPP-----GHPLPLGLHDTAR--FVLPRPPTISNPPPLISSAKHPSVLERQI	1122
Db	1009	LPEQARAAPATGSOQLILGHAPAYADVAQAQVTVVPVPAAVLSPL-----PEVL---	1058
Qy	1123	GAISQGMVSQVLHVPYSEHAKAPGVFTMGLPLPMDPKKLAPFGGVKQEOQLSPRQAGPPE	1182
Db	1059	-----LPAAPEL-----LPQFPSSLATVSASVQ-----	1081
Qy	1183	SLGVPTQAEASVLRGTALGVPGG-SITKGIPSTRFVSDSAITYRGSITHGTADVLKYG	1241
Db	1082	---SVPT---QTATLLPPANPPLPGFGFIASPCPTVQLTVPEVQOEAS-----	1124
Qy	1242	TITRIIGEDSPRLDRGREDSLPKGHVIEYEGKHVLSYEGGMSVTQCSKEDGRSSSGPP	1301
Db	1125	-----QDKPPGLPQSCF-----SY-GGSDVT-----SG--	1146
Qy	1302	HETAAPKRTYDMMEGRVYGRAISSAIEGLMGRAIPPERHSPHHLKEQHIRGSTIQGIPR	1361
Db	1147	-----KELSDSCBG--AFGGGRLEGRAAR-----KXH-----	1171
Qy	1362	SYVEAQEDYLREAKLKLKREGTPPPPPSRDLTAYKYQALGPLKPAHEGLVATVKEA	1421
Db	1172	-----RRSTRASRQERASRP-----RLT-----ILNVCNT	1197
Qy	1422	GRSITHEIPREELRTHPELPAPLPLKEGSIITQGTPLKYDTGAS-----TTGSKKHDRVSL	1476
Db	1198	GDKWVECOLETHNH-----KWTFKFDLGDADPADDEIATVMWEDFP--I	1238
Qy	1477	IGSPGRTFPPVPHLP-DVMADARALERACYEBSLKSREGTA-----SSSG---GSJAR---	1524
Db	1239	LOAERETP--IEQMKDVMKAEMLSEDTDADRQSDPCTSPPHLSTCGLGTGGEESRQSA	1296
Qy	1525	GAPVI---VPBLGHPGROSPLTYEDHGAPPAGFAGHLPRGSFVTVWRETPPLQEGSLSSSKASQ	1581
Db	1297	NAPVYQNVLTGKRWFITICPVABHAPAEA---PESSP-----PLP-----	1334
Qy	1582	DRKLTSPRETAKSPHSTVPEHHPHPIPSYEHLLRGVSGVDLYEXSHIPLAFDPTSI PRGI	1641
Db	1335	---JSSUPPEASQFCRGL-----TLPCULPMRRAACGAVFUSLFSAE-----	1373
Qy	1642	PLDAAAAYTLPRHLAPNETYPHLYPPVPLIRGYPDTAALENRQTLINDIYITSSQQMHNTAT	1701

Db	1374	----	SAQSQPPDSAP-----YKQLSSKEQPS-----FLASQJL-----	1404
Qy	1702	AMAQRMLRGLSPRESSALNYAAGPRGIIDLISQV- PHLPLVLVPPTP-----GTPATAMD	1756	
Db	1405	-----LSQAQSPNPPCAPPAPLAPSPPPVTALP	1432	
Qy	1757	RLAYLP---TAPOPFSRHSSSPLSGGGPHLTLPKPTTTSSSERRDRDRDRDREKS	1813	
Db	1433	QDGAAPATSTMPFASGTASQAG-GFGTPOGLTSETLSQPLAE-----THEAP	1480	
Qy	1814	ILTSITTVEHAPIWRCTEQSSSGSSGGGGSSSRPASHSHAHOSHPISPRTODAL-Q	1872	
Db	1481	LAVOPLVWGLAPC-TPAPEAASTRD-----ASAPREPLPPAPRPSHSGTPOPALQG	1532	
Qy	1873	QRPSVLHNTGMKGIIITA-----VBP SKPTVLTRSTSTSPVRPAATFFPATHCP LG	1922	
Db	1533	PAPLLPAAVGAVSLATSOLPSPPLGFTVPQPPSALESDGEGPP--PRVGFDVSTI----	1586	
Qy	1923	GTLDGVVPTLMBEVLLPKCAPRVARPERPADTGHAFLAKPPARGGLEPASSPSKGESEPR	1982	
Db	1587	KSLDEKLRTLTYQEHVPTSASAGTP-----VEVGDRDFTELEPLR-GDQPRESEVCGGDLAL	1641	
Qy	1983	PLYP--PVSGHATIA R--TPAKNLAPHASDPDPAPASADPHREKTOKSPFSIOELEL	2038	
Db	1642	PPYPKEAVSGRVLOPOLPVEKSELATRGAVMGQTSSS-----MTESSPRSM-----	1689	
Qy	2039	RSLYGHGSS---YSPEGVEPVSPVSPSLTHDKGLPKHLELDKSHLEGELRPKPQCPVK	2095	
Db	1690	--LOYDRDHQVADSHVPSVDQVPAFVR----PARVEPTRDDGCEAGESAEPPSPD	1743	
Qy	2096	L---GGEEAAH-----LPHLRPLPESQSSPLLQTA PGVKGHORVWTLAGHISEVIT	2144	
Db	1744	MGTVGGQASHPOTLGARALGS PRKRPEQQDVSSP-----AKTVGRFSVVST	1789	
Qy	2145	OD-YTRHHPOOLSADLPAPLYSPFGASC P--VLDLRRPPSDLYLPDPDHCA PARGSPHSE	2201	
Db	1790	QDEWTLASPHSLRYSA PPDDVYLDEAFSPDPVKLA VRRAQTASSI-EVGVGE PVSSDSGD E	1848	
Qy	2202	GGKRSPEPNKTSVL--GGGBDIGEPVSPBGMTEPGHSRAVYPLLRYRDEQETEPSRMGS	2259	
Db	1849	GPRARPVQXQSLPVSGSVAGDFVKATAFIQRPSPAGSL-----GPET-PSEVGM	1899	
Qy	2260	KSPCNTSQPPAFFSKL-----TESNSAMVKSXQEIINKK-----LNTHNRNEPEY	2304	
Db	1900	KVP--TISVTSFHQSQSYISSDNDELEDADIKKELQSLREKHLKEISELSQSQKQBIEA	1957	
Qy	2305	NISQ-----PGTEIFNMPAITG LMYRSQAOEHASTNWGLEAIRK-----ALMGK	2353	
Db	1958	LYRLRGKPLPNVGFHTAPTPTGRRKTKSKK-----LKAGKLLNPLVROLKVVASSTGH	2012	
Qy	2354	YDQWEESPPLSANAFNP LNASLPAAMPITTAADGRSDHT-----LTS	2396	
Db	2013	LADSRGPP-----AKDPAQASVGLTA--DSTGLSGKAVQTQPCSVRASLSSDICSLAS	2066	
Qy	2397	PGGGKA-----KVGSRPSSR-----KAKSPA	2418	
Db	2067	DGGGARQQGWVYHPTSERV TYKSSSKFRPFLSGPVSIWSALKRCLCGKEHSSRSST	2126	
Qy	2419	PGLASGRDPPSVSVHSEGDC-----NRTPLTN RV-----WEDRPSSAGSTFPYNPL-	2467	
Db	2127	SSLAPGEPGPQPALHVQAQVNNSNNKGTFTDDLHKLVDEWTSKTVGAQAQLKPTLNQLK	2186	
Qy	2468	-----IMRLQAG-----VNASPPP-----PGLPAGSGPLA	2492	
Db	2187	QTOKLQDMEAQAGWAAPGEARAMTAPRAGVMPRLPPAPGPLS	2229	
RESULT 60				
Q9VC00 PRELIMINARY; PRT; 2768 AA.				
ID	Q9VC00			
AC	Q9VC00;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			

Db 1528 PLAGDEESNLPKLPQDIFEEAPVAVTTAAPSXDDGGEQKPEVEEKEPDIQGXPIEDET 1587
Qy 1573 -----SLSSSKASOORKLSTPREIAKSPHSTVPEHHPHPIPSYEHLLRGVSGVDLYRSH 1627
Db 1588 STPTSSNEIPESDRATTIAPSKEESEPSTGAPTQDEPAEPSTDAPEDESKEETPESE 1647
Qy 1628 IPLAFDPTSIPIRGIPLOAAAAAYILPRHAPNTPTPHLYPYLIRGYPDPTAALENRQTIIN 1687
Db 1648 VP-----ITVAGEKIPSS---ITPDEEPTATAPVAKP-----DEDVEKETSTRIPT 1694
Qy 1688 DYITSQQMHNTAT-----AMQADMLRGLSPRESSLALNAAAGRGIIDLSQVPHLP- 1741
Db 1695 DAPASSEDENSSTDQIPSEVPSEKPPETPAQTPAQTBEGDI-VGATAAPTTSDEVPVQRLPE 1753
Qy 1742 -VLVP-PTPG-----TPATMDRLAYLTAQPPFSRRSSSPPLSPGCGTHLTKP 1788
Db 1754 EVLAEIPOPSTETGKQODETTAAPSIDRKE--PYVTEIDBEATTVAPISEKD-----EKP 1807
Qy 1789 TTTSSSERDRDRDRDRERKSIILSTTTTVEHAPIWRPQTQS--SGSSSGSGGGGGS 1847
Db 1808 TEEKPVEQKTPGEPSSEKEKEPIEQDVST--EGPV---STASEAGSTESS----- 1856
Qy 1848 SSRPASHAHQHSPISPTQDALQORPSVLHNTGMKGIITAVPSKPTVLRSTSTSPV 1907
Db 1857 -----BEVKPSTEGEVAEKP-----EDKQP-----SSTAQ 1881
Qy 1908 RPAATFP-ATHCPILGGTLDGVYTLMEPV-----LLPK-----EAPRV--A 1946
Db 1882 APVTIPEISTELP---AQDGKPTSEAPVDSDEDTSAPEDEKIPSVSGEEVEGPEVTTA 1938
Qy 1947 RPERPRADTGHAFKAPARGCLPASP-----PSKGSPPRLVPPVPSGHAT 1993
Db 1939 SPQAAEDE-----LKTPAES--EPSSDTKVPETEKPEDETKADETPESVTQVSDVAT 1991
Qy 1994 IARTPA-----KNLAPHASDPDP-----APPASDPHRKTKOSKPSIQELE--- 2037
Db 1992 STSAPVAGGDIKDEQATFASPEEEIEKTIAPAAEIPQ-----SEKPEVDEQEVESGT 2047
Qy 2038 ----LRISGYHGSYSPEGVSPVSPSPLTHDKGL-----PKHLEL----- 2077
Db 2048 KATPAESDQPIDEIAPATSGPIDEASTAATKESITTVASAAAPVHDDBIKDVTTTQP 2107
Qy 2078 --DKSHL-----EGELRPKQPGVPLKGLGEEAHLPHRLPLPSQP 2114
Db 2108 VADEKEVAAPQDETKTSIDVSTDSTPAQDDEKQDKTEAPV-----APTIVSSP 2155
Qy 2115 -----SSPFLQAPGVKHQV-----TLAQHISEVITQDYTRHHHQQLSAPL-- 2159
Db 2156 TADSAADSSTTVEVPSPVEIDTKPMDDIMSQTIAHTADGAASTSTEDEDQ---APVTV 2212
Qy 2160 -----PAPLYS-----FPGASCPLDLRPPSDLYLPPDHGAPA 2194
Db 2213 SPQDAKTPVPAPQDSKTPSSBAPQDAEIPATATPLDNDKIPAT--VAPQDDDGVA 2270
Qy 2195 RGSHPSECKRSPBNKTSVLGGGEDGI-----EPVSPPEGMTEPGHRSAPV 2241
Db 2271 TAAPLDE-----DKIQTTAAPLDEEKIPSTAAPLDDEKIPAPVSPVDFVPEPSSEKPAV 2324
Qy 2242 YPLLYRDEGEQTEPS-----RMGSKSPGNTSQPPAPFSPKLTENSAMVK 2284
Db 2325 SE--YDGESETEPPVHDVETSTDEPTSDAKLKPTTAPATPSESPPATEAEIVPETAPEL 2382
Qy 2285 SK-----KOEINKLANTHNRPEYNIQSP--GTEIFNMPAITGTLMTYRSQAVOE 2334
Db 2383 KEVPEKATEQPELEKETPEKATQOPELEKETPEKATE---QPEL-----EKETPE 2430
Qy 2335 HASTNMGLEAIIRKALMGKYDQWEEESPPLSANAFNPLNASLPAAMPITAADGRSDHLL 2394
Db 2431 KATEQPELEKEVT-----DKATEQSPESVDEKTP-----EPVWKPSLDSTEDEBSV 2477
Qy 2395 TSPGGG-----KAKVGRPSRKAQSP--APGLASGDRPPSVSVSHSEGD-----CNRR 2442
Db 2478 ESBESADKDKNKRETEEDTDKGHEEPVPAVSEIPOPSEAVPTTGHPLFPHLASSTT 2537

Qy 2443 TP--LITRWVED-----RPSSAGSTPPFPYNPLIMRLQAGVMAS-----PPPPGLPA 2486
Db 2538 TTPAVDDRVGEEDENTTVKLSSTTSTTSTESPTVSAPTSTTTVASQQQPPITPPYG--- 2594
Qy 2487 GSGPLAGPHHA-----WDEE 2501
Db 2595 -----HAPEYDEYDEE 2606
RESULT 61
Q8CF91
ID Q8CF91 PRELIMINARY; PRT; 4969 AA.
AC Q8CF91;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Piccolo.
GN PCLO.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22384373; PubMed=12401793;
RA Fujimoto K., Shibaaki T., Yokoi N., Kashima Y., Matsumoto M.,
RA Sasaki T., Tajima N., Iwanaga T., Seino S.;
RT "Piccolo, a Ca2+ Sensor in Pancreatic beta-Cells";
RL J. Biol. Chem. 277:50497-50502(2002).
DR EMBL; AB083478; BAC53724.1; .
DR PIR; PT0546; PT0698.
DR PIR; PT0644; PT0644.
DR MGD; MGI:1349390; Pclo.
DR GO; GO:0045202; C:synaptic junction; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR008899; Znf_piccolo.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF05715; Zf_piccolo; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR PROSITE; PS0106; PDZ; 1.
SQ SEQUENCE 4969 AA; 540677 MW; 2A3F9D372BE8B2D CRC64;
Query Match 3.4%; Score 448.5; DB 11; Length 4969;
Best Local Similarity 19.8%; Pred. No. 1.4e-12;
Matches 397; Conservative 226; Mismatches 684; Indels 695; Gaps 95;
Qy 688 AAASEEAAFPVVDDEMEASGVSGNEEEMVEEAEALHASGNEVPRGECSPATVNNSSD 747
Db 23 AGGSGSALHFGIPAGMEADLSQLS--EERERQIAAVMSRAQGLPKG--SVPAASAAES-- 75
Qy 748 TESTPSPHTEAAKDTGONGKPKPATLGADGPPGPPPTPRR-TSRAPTEPTPASEATCAP 806
Db 76 ----PSMHRKQELSSQ-AFQQP-----GKPPDGRPPQHLGSKSRITDTFRSE---Q 120
Qy 807 TTPPAPPSP-----SAPPVVPK-----E 825
Db 121 KLPGRSESTISLKESKSRDTFKEYKSMMPGPFSDVNPPLSAVSVVKNKFPNPPDLISDE 180
Qy 826 EKEBETAAAPPVEGEBOK-----PPAABELAVDTGKAEE-----PVKSECTEE 869
Db 181 AVQEBETTKQKVAQDKQKSEGITKPSLIQQPSPLIPKQQPGKGEVTPQDIPSKSVSSQ 240
Qy 870 AEE-----GPAKGDAAEAATAEGALKAEEKGGSGRATT-----AKSSGAPQSDSS 918
Db 241 AEKTKPAQPGTAKPSQSPQATPAQA-KPVAQQPGPAKATVQQFGPAKSPAGTQGS 299


```
Qy 2315 NMPAI---TGGLMYRSQAOVHASTNMGL-EAIRKALMGKYDOWEESPLSANFNP 2370
Dy 1456 NSPVARRKRRTSIGSSSEYKQDSQSGDEDFIRKQII-----EMSADE--- 1502
Qy 2371 LNASASIPAAPIITAADGRSDHTLITSPGGGKAKVSGRPSRKAAPSAPGLASGDRPPSV 2430
Dy 1503 -DASGS-----EDEFIRSQLKEIGGVTEQKRETKGKGSFA---GKHRLTRK 1549
Qy 2431 SSVHSEGDGNCRRRTPLTNRWED 2452
Dy 1550 SSTSFDDAGR-----HSWHD 1566

RESULT 63
ID Q86WG6 PRELIMINARY; PRT; 2948 AA.
AC Q86WG6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Transforming acidic coiled coil 2, long isoform.
GN TACC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22508191; PubMed=12620397;
RA Lauffart B., Gangisetty O., Still I.H.;
RT "Molecular cloning, genomic structure and interactions of the putative
RT breast tumor suppressor TACC2.";
RL Genomics 81:192-201(2003)
DR EMBL; AF528099; AAC62630.1; --
SQ SEQUENCE 2948 AA; 309400 MW; 46711F3CD0ADCDC CRC64;

Query Match 3.4%; Score 446; DB 4; Length 2948;
Best Local Similarity 19.9%; Pred. No. 9.3e-13;
Matches 499; Conservative 229; Mismatches 865; Indels 920; Gaps 114;

Qy 678 NARRKKKAPAAAEAAFPVVEDEMEASVSGNEEMVEEABAL-----HAGNEVP 732
Dy 5 NSTSDNRTLSAQTPRSAQPP-----GNSQNIKKQQTDPGSPDHRDASSIG 51
Qy 733 RGECSGPATVNSSDT--ESIPSPH-TEAAKD-TGONGPK-----PPATLGADGPPPGP 782
Dy 52 SVGLGGFCTAESASLDDCLVSPVTEPRKDPOQARGPESGLLPPSPPOERHPPSSM 111
Qy 783 P---TPPRRTSRAPTEPTPASEATCAPTPPPAPPSPAPPVVPVPEKE-----BETAAP 835
Dy 112 PFAECPPEGCLASPAAPEDGPQTSPPREPAFAPAGDIAAFAERDSSTPYQEI AAVP 171
Qy 836 -----PVERGE-----RQKP-----PAAEL-AVDTKABEPVKSECTEAAE 872
Dy 172 SAGREROPKEKGKSSFSFGIDQSPGMSVPLREPMKAPLCGEGDQPGGFESQEEAA 231
Qy 873 G---PAKGKDA-----EAAEATAG-----ALKAKEGG 899
Dy 232 GGFPPAESRQGVASVQVTPPEAPAAQQTESAVLEKSPKMAPIPODPAPRASDRERG 291
Qy 900 SGRA-----TTAKSGAPO-----DSSSSATC----- 921
Dy 292 QGEAPPOYLTDLFLRACHLPRNSGAPAEVNAASQESCCQPVGAYLPHAEPLWCLP 351
Qy 922 SADEVDAEGGDKNRL-----LSRPSLLT 946
Dy 352 SPALVPEAGSGKEALDTIDVQGHPTGMRTKPNQVVCVAAAGQOEGGLPVSPESILT 411
Qy 947 PT----- 948
Dy 412 PTEBAHPASSIASFPAAQIPIAVEEPGSSRESVSKAGMPVSADAKEVDAGLVGLERQ 471
```

```
Qy 949 -----GDPRANASPO-----KPLDLKOLKORAAAIPPIQVTK---VHE----- 983
Dy 472 VSDLGSGXHEPGEQGVGPVAPSPQERGEHLNTEQSHVQVGPVPPPLPKQESHEVQCAP 531
Qy 984 PPREDAAPTKPA--PPAPPPQNLQPSDAP-----QOPGSSPRGSRSPAPP----- 1029
Dy 532 PPPLPKAPSSARGPPTTGAKVHEDSTSPAVAKESRSRSPGDSPPGKEEAPEDPDGDDP 591
Qy 1030 -----ADKEFAAABAQKLPDGPCCWTSGLPPVPVPREVIKASPHAPD---PSAFSYAPP- 1080
Dy 592 GNLOGEDSQAFSSKR-----DP-----EVCKDELSKPSDDAESRHFSSHAQPPR 637
Qy 1081 -----GH-----PLPLGLHDTARVPLPRPPTISNPPPLISSAKPSVLER 1120
Dy 638 KKGAGHTDGPHSQTAEADASGLPHKLGEED---PVL-----PPVFDGAGEPTPE- 684
Qy 1121 QIGAISQMSVOLHVPYSEHAKAPVPTMGLPLPMDPKKLAPFGVGVKQEQLSRGOAGP 1180
Dy 685 --GAIWEGSLQPKCPDTLQSRGLGRMESPLTLESEKSDPPP---TPVAEVAKQAGG- 738
Qy 1181 PESLGVPTAQEASVLRGTALGSPVGGSIITKIPSTRVPSDSAI--TYRGSITHGTADVL 1238
Dy 739 -----ESTLEIKWGS CDGEGLLTSPDQPRGPACDASRQEFHAGVPHPPQGENL 787
Qy 1239 YK--GTITRIIGEDSPRLDRGRDSDLPKHGVIYEGKKHVLVSYEGGMSVTQCKEDG-- 1294
Dy 788 AADLGLTALILDQD-----QOGIPSCPEGWIRGAASEWPLLSSEKHFQPSQAQPTSIF 842
Qy 1295 ---RSSGPPHETAAPKRTYDMMGRVRAISSASIEGLMGRATPPE---RHSPHLKE 1347
Dy 843 DVLKEQAOQPPEN---GKETSPSHFGKQDQADSSQIH-----VPVEPEQDNNLPTHGQ 893
Qy 1348 QHTRIGSITQIPRSYVE-----AQDDYLRRRAKL-----LKRGTPTPP----- 1386
Dy 894 EQALGSELQQLPKGTLSDTPTSSPTDNWESSLTERSELSAPTRQKLPALKEKPEGAC 953
Qy 1387 -----PPPSRDLTEAYK-----TQALGPLKPAHEGLVATVKEAGRSIHE 1427
Dy 954 GDQGSRRVSPPAADVLDKDFSLAGNFSRKETCCTCGPNKSQA-----LADALEG-SQHE 1008
Qy 1428 IPRELPH-----TPELPLAPRLKEGSIQTGTPLKYDT----- 1461
Dy 1009 EACO--RHPGASEAADGCSPLWGLSKREMASGNTGEAPPQCPDVSALLDAVPCLPALAPA 1066
Qy 1462 --GASTTGSKKHVRSGLISPGRTFPVPHVLDVNVADARALACRACVEESLKRPGT----- 1514
Dy 1067 SPGVPTQDAPETACDETQEGROQVPAPQOKM-ECEATSDAESPKLLASFPSSAGEGG 1125
Qy 1515 ---ASSGGSIARGAPVIVPELKGPROSPLTYEDHGAPFAGHLPRGSPVTWREPTFRLQE 1571
Dy 1126 EAGAAETGGSAGADPGKQQAPEKPEATLS-----CGLLQTEHCLTSGE-----E 1171
Qy 1572 GSLSSSKAS-----QDRKLSTPRETAKSPHSTV-----PEH 1603
Dy 1172 ASTSALRESQAEHPMASCQDALLPA--RELGGIPRSTMDFTHQAVPDPKELLLSGPE 1229
Qy 1604 HPHPISEVHEL---LRGV---SGVDLYRSHIPLAFDPTSI PRGIP---LDAAAAAYLPRH 1654
Dy 1230 VAAPDTYLVHDSAAQGAEDSGVKAVSSADPRAPGESPCVGPFPPLALENAASL---K 1285
Qy 1655 LAPNTPYPHLYP-----SPRESSLA-----LNYAAGPRGIIDLSQVPHLPVLVLP 1676
Dy 1286 LFAGSLAPLQPGAAGGEPAPVAVAGSSGPKARTTEGVPVDSMPCLDRMPLAKGQATGEE 1345
Qy 1677 -----ALENROTIINDYITSQMHNTATAMQADMLRGL--- 1713
Dy 1346 KAATAPGAGAKASGEGMAGDAAGTEGSMERMERFSPDQKQGTSGGVDTSSEQATLTGF 1405
Qy 1714 -----SPRESSLA-----LNYAAGPRGIIDLSQVPHLPVLVLP 1745
Dy 1406 PDFREHTAKIFEKPVILGALATPGKAGAGSAVKDLTRPLGPEKLLD--GPPGVDVTL 1463
Qy 1746 PTPGTATAMDLAYLPTAQPFSSSRSSSLSPG--GPTHLT-----K 1787
```

Db 1464 PAP--PARLQVEKKQQLAGEAIEISHLALQDPASDKLLGPAGLTWERNLPGAGVGKEMAGV 1521
Qy 1788 PTTTSSSER-----ERDRERDRDREREKSLTSTTTTVEHAPIWRPQTEQSSG 1836
Db 1522 PPTLREDEPRPGGAAMPGLGQAYSLQERSQELASGLPSAATQELPVERAAAFV-- 1579
Qy 1837 SSGSGGGGGSSSR--PASHSHAHQSPISPTQDALQORPSVLHNTGMKGIIITAVEPSKP 1895
Db 1580 APHSHGEAVAQDIRPSGKHQHOETACDSHPHGEDG-----PCDFAHTGVPG----- 1625
Qy 1896 TVLSTSTSPVRPAATFPFPAATHCP-----LGG-----TL 1925
Db 1626 HVPRSTCAPSQREVLTVPANSESPFWLTLTGERRPGVTAGILEMRNALGNQSTPAPPT 1685
Qy 1926 DGUYPTLMEP-----VLJPKEAPRVARPERPRADTGCHAF-----LAKPPA 1965
Db 1686 GEVADTTPLEPGKVAGAAAGEAGDITLTAETQACASGDLPEAGTTRTFVSVAGDLVLP-- 1743
Qy 1966 RSGLEPASSPSKSGSEPRPLVPPVGHATI-ARTPAKNLAPHASGDP-PAPPASASDPHR 2023
Db 1744 GSCQDPACSDK-----APGMEGTALHGDSPAR---PQAKEQGPGRPIFAGD--- 1789
Qy 2024 EKTQSKPFSIQEELRSLGYHSSYPGVEPVSPSLTHDKPLKLEELDKSHLE 2083
Db 1790 -----GKVCVSSPPPEPDTHD---PK-LQLHAPAEELH 1817
Qy 2084 GELRPKQPGPKVL-----GGEAAHLPLRLPESOPSSPLLQ 2122
Db 1818 TDRESPPRGPMPLSPVKPADRVMDKVTSDTGAEGT-----ESSPVADDDIQQ 1868
Qy 2123 APGVKGHQRVTLAQHISEVITQDTRHHPQOLS-----APLP-----APL--- 2163
Db 1869 AAPADLESPTLAASYHCDVVGQVSTDLIAQISPAHAAGLPSAAEHVSPAPAGDR 1928
Qy 2164 --YSPFGASCPLDL--RRPPSDLYLPPDHGAPARGSPHSGGRKSPENPKNTSLVGGGED 2220
Db 1929 VEASTPSCPDPADKLSRSSDSEAFETPSTTPVKAPP-----APPPPPPEVPEPEV 1981
Qy 2221 GIEPVSPGEGTEGHRSAYVLLYLDGEOTESPMSGKS---PGNTSQPPAF-FSKLT 2276
Db 1982 STQP--PPE-----EPG-----CGSETVPDPGRSDSVESGSPFRPFSHSAVF 2024
Qy 2277 ESNAMVSKK-----QEINKLNTNRNEPEYNISQPGTEIFNMPAITGT 2322
Db 2025 DEDKPIASSGTYNLDFONIELVDVTFQLEPRASDAKQEGKVNTRRKSTD--SVPIKST 2082
Qy 2323 GLMYRQAVQ-----EHASTNMGLEAIRKALMGKYDQWEEPSPLSANFNPLNASAS-- 2376
Db 2083 ---LSRSLSLQASDFDGAASSGNPEAV-----ALAPDAYSTGSSASST 2123
Qy 2377 -----LPAAM-----PITAADGRSDHTLTSPPGGGKAKVSGRPSRRKAKS 2416
Db 2124 LKRTKPRPPSLKKKQTKTETTPVKTEQOEDEESLVPSGENLASSET-KTESAKTEG 2182
Qy 2417 PAPGL-----ASGDRPPSVSS--VHSEGDGNRRRT-PLTN--RV 2449
Db 2183 PSPALLEETPLEPAVGPKAACPLDSESAEGVVVPASGGRVQNSPPVGRKTLPLTTAPEA 2242
Qy 2450 WEDRPS--SAGSTPPFPYPLNLRMQGV-----MASPPPPGLPAGSGPLA 2492
Db 2243 GEVTPSDSGGQEDSPAKGLSVRLRFDYSEDKSSWDNQENPPPTKIGKPKVA 2295

RESULT 64
Q81SF6
ID Q81SF6
AC Q81SF6
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 23, Last annotation update)
GN 2Mda_2 protein.
ISOF.

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Mutations in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions.";
RL J. Mol. Biol. 323:533-549 (2002).
DR EMBL; AY130758; AAN61518.1; -.
DR PIR; E89066; E89066.
DR PIR; T33247; T33247.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR001023; Hsp70.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_kinase.
DR Pfam; PF00041; fn3; 11.
DR Pfam; PF00047; ig; 43.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000089; Hsp70; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00060; FN3; 11.
DR SMART; SM00408; IGC2; 37.
DR SMART; SM00406; IGV; 4.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; Ty_Kc; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00835; IG_LIKE; 38.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 18519 AA; 2051869 MW; 6A8441C5D0BA7729 CRC64;
Query Match 3.4%; Score 445; DB 5; Length 18519;
Best Local Similarity 19.0%; Pred. No. 1,1e-11;
Matches 475; Conservative 334; Mismatches 1017; Indels 672; Gaps 109;
Qy 51 SHLSPGSIQ-----PQRRPSSLSEFQCNERSQELHLRPESHSHVLPGLKSEMEFIE 104
Db 4934 SEVQQAIIQKQKQVVPVETSAPT-----EPTVE-----LKPVESKETSEV--QQVEIIE 4982
Qy 105 SKRPRLELLPDLRLPS-PLLATQAPAGSEDLTKDRSLTGKLEVPSPSPPTDPPELVL 163
Db 4983 QKQVVPETSAPTVEPTVEPTVEKLPVESKETSEVQQAIIQKQKQVVPVETSAPTVEPTVEKL 5042
Qy 164 PPRLSKE-----ELIQNMDRVDREIT--MVEQIISKLK-----KKQQLSEEA 204
Db 5043 KPVESKETSEVQQAIIQKQKQVVPVETSAPTVEPTVEKLPVESKETSEVQQAIIQKQK 5102
Qy 205 AKPEPEKPVSP-----PIESKHSRLVQIIDYENKKAFAAHRILEGLGPGVQLPLYN 258
Db 5103 VVPVETSAPTVEPTVEPTVEKLPVESKETSEVQQAIIQKQKQVVPVETSAPTVEPTVE---KH 5159

Query Match 3.4%; Score 445; DB 5; Length 18534;
Best Local Similarity 19.0%; Pred. No. 1.1e-11;

Matches	475; Conservative	334; Mismatches	1017; Indels	672; Gaps	109
Qy	51	SHLSPGSIQ-----PQRRRP	51	SHLSPGSIQ-----PQRRRP	51
Db	4934	SEVQQAIIHQDVPVPETSA	4934	SEVQQAIIHQDVPVPETSA	4934
Qy	105	SKPRLELLPDP	105	SKPRLELLPDP	105
Db	4983	QKQDVPVPETSA	4983	QKQDVPVPETSA	4983
Qy	164	PPRLSKE-----ELION	164	PPRLSKE-----ELION	164
Db	5043	KPVESKETSEVQV	5043	KPVESKETSEVQV	5043
Qy	205	AKPEPEKVPSP-----PIES	205	AKPEPEKVPSP-----PIES	205
Db	5103	VPVPETSA	5103	VPVPETSA	5103
Qy	259	QPSDTRQYHENIKINQAMR	259	QPSDTRQYHENIKINQAMR	259
Db	5160	APVESKETSE-----	5160	APVESKETSE-----	5160
Qy	319	N--PRRAKESKVREY	319	N--PRRAKESKVREY	319
Db	5186	TSAPTVEPTVEK	5186	TSAPTVEPTVEK	5186
Qy	377	DGLSEONLEKQ	377	DGLSEONLEKQ	377
Db	5230	--JTEVE--VELF	5230	--JTEVE--VELF	5230
Qy	434	EQBKETPREKPMQ	434	EQBKETPREKPMQ	434
Db	5284	VESKETSEVE-----	5284	VESKETSEVE-----	5284
Qy	484	RRYRRRGKSG	484	RRYRRRGKSG	484
Db	5324	-----ESKETSEVQQA	5324	-----ESKETSEVQQA	5324
Qy	525	--KEKAEEKEE	525	--KEKAEEKEE	525
Db	5376	VEQKDVTCSEI	5376	VEQKDVTCSEI	5376
Qy	579	KGRITRSMANEANSE	579	KGRITRSMANEANSE	579
Db	5436	EPITVEK	5436	EPITVEK	5436
Qy	639	RMVGSKTVCCKN	639	RMVGSKTVCCKN	639
Db	5482	KSVESKETSEVQ-----	5482	KSVESKETSEVQ-----	5482
Qy	687	--PAASEEAAFP-----	687	--PAASEEAAFP-----	687
Db	5531	VQPAIVEQKV	5531	VQPAIVEQKV	5531
Qy	739	PA---TVNNSSD	739	PA---TVNNSSD	739
Db	5589	PTVEPTVEK	5589	PTVEPTVEK	5589
Qy	796	PTPASEA-----	796	PTPASEA-----	796
Db	5642	SKETSEVEPA	5642	SKETSEVEPA	5642
Qy	836	PVEGEEOQPPA	836	PVEGEEOQPPA	836
Db	5702	--ETSAPTVE	5702	--ETSAPTVE	5702
Qy	888	EGALKAEKKEGGS---	888	EGALKAEKKEGGS---	888
Db	5752	EKLAPVESKETSEVQ	5752	EKLAPVESKETSEVQ	5752
Qy	934	KNLLSPRSL	934	KNLLSPRSL	934
Db	5812	QKDV--VPVPETSA	5812	QKDV--VPVPETSA	5812

Db 78 QAPPPETPSQSQOQLOPQVLDAPAHKLDVALENGISYTLNSPPRPAISPLTD 137
Qy 836 PVEGEBOQP---PAABEL-----AVDTGKABEVPKSECTEABEGPAKAGKDAEATAE 888
Db 138 PVPDPRTREPSVQPLESELETHAADQPKTINGIAHSNPNRPNNTDLKGRD--TPQVTSV 195
Qy 889 GALLAEKKEGSGRATTAKSSGAGQDSDDSSATCSADEVDEAEGDKNRLLSRPSLLTPT 948
Db 196 GA-TAGATVGGAAATTVFVGN--STNSTSHFSGLAWDGTTP-----IAPFNTQAPPA 243
Qy 949 GDPRANAS-PQKPLDLKQ-----LQRAAAALPPI----- 976
Db 244 VDGWANDWFFSTWDLLOSQSGHILMSFLOSLSVTTQQOQLOQIPLPPLPNTVRPSQVSLA 303
Qy 977 -----QVTKVH-----EPREDAAATKP----- 994
Db 304 QMQLDOANATHPLHLTAQGGIQPPAAAAATAADASLESFARLEFADGVFQMTYALIIG 363
Qy 995 -----NLOPES 1009
Db 364 RDQRAWLAKKEERRAEQYQBLKDEYEAQGLTPPAPPSDEDARRFSKSYISEEGNLPES 423
Qy 1010 DAPQPGSSPRGKSRSPAPPADK-----EAF-AEAQKLPDPPCWTGSLPFPVPP 1059
Db 424 DDEDEBCRP--SDNRAPKKRKNMGVSLPVDGFEAESLSMSADQT--VDDGKGPSVN 478
Qy 1060 REVIKASPHADPDSAFSAVPGHPLP-LGLHDTARPVLPRPPTISNPP----- 1106
Db 479 RQYVSHTPGAAAVNLSALRPSFHTFPLGHSFGFNIAAKTKAISREHLKTIQFNSQAGVF 538
Qy 1107 ---PLISSA-----KHPSVLERQ-----IGAISQGMS-VOLHVP--- 1136
Db 539 EAIPLHKNFGFCEDVHYSHDKVILKSGDRLOVKQVDFVFIINGVAEGKTGAEEYEPETP 598
Qy 1137 ---YSEHAK-----APVGEVTMGLPLPMDPKLAPFPQKQQLSPRQAG 1179
Db 599 ARRYSEGCKEMSFDPESITHVDRRSTSPEDNEVMSVDPDR-----DDSQSELS---EP 649
Qy 1180 PPESLGVPTAOB---ASVLRTGA-----LGSVPGGSITKG 1211
Db 650 PDEDILLDPADPHVMTVEKEEAEDDDDEARSQSQSVKPRPPEPFDPMSSVPLMDPPKK 709
Qy 1212 IPSTRVSDSAITVRSITHTGTPADVLYKGT-----ITRIIG-----EDSPS 1253
Db 710 RQGRPPXNGIMSKEERLRKQAMELAKNQPPQPCGPPVKXKVGPRKHPLEDAPD 769
Qy 1254 RLDGREDSPKGVHVIYEGKKGHVLSEYEGGMSVTQCKEDGRSS--SGPPHETAAPKRTYD 1312
Db 770 RPEKRYKPRK-----NGEEDGASDAE-----KTIKEKREKPKTPPLELRREDYTEE 818
Qy 1313 MME-----GRVGRAISSASTEGLMGRAPPERHSPHLKQHHIRGSIOTGIPRS----- 1362
Db 819 QLOQKNKNGYGLIDVLSAAPDGLTKQI-----YKRIQLKYPFYFVNDTKGWSVRHN 874
Qy 1363 -----VVEAQDYLREAKLL-----KREGTPPPPPPS--RDLTEAKTQALGPKLKP 1409
Db 875 LIGNDAFKNEETHLWSRVPIDIDAGKKRKAPSFDHASSLHNFQCHVAPQM-----P 928
Qy 1410 AHEGL-----VATVKEAGRSIHEIPREELRH-----TPELPLAP 1443
Db 929 PHPGVMYHGBGVQSVHPGTVGQRQSVYTTGQPGASQH--PPQHLQTPQPGVPPQP--P 984
Qy 1444 RPLKSGSITQGTPLKYDVGASTTSGKKHDVRSLLSGRGTFF-----PPV----- 1487
Db 985 REAYQAAQTSPPAQ-----AQYVGTPPTAARQMSGTPAATYSSPVSRPMPMTVAASGA 1040
Qy 1498 -----HPLDWM--ADARALERACYBESLKSRRPGTASSSGSSTARGA----- 1526
Db 1041 TPHSWARQHSILFVSGFSGANGIPRV-----NPPATANTGVVAGGAPPAQQAPTTAN 1093
Qy 1527 -----PVIVPEL-----QKPRQSPLTYEDHGAPFAGHLIP-----RGSVPTWREBTP 1567

Db 1094 TTAPVRLNVPVIAPELISWLESFKVTVEKLEYIQKTSKFPQILAMSVINRGLKLTTKSMIP 1153
Qy 1568 -----RLQ-----EGSLSSSKAS----- 1580
Db 1154 DEESLELVURVPEERIQTGTSKSLDPLDQLLQTLTTFKATMVSTLEAKLDSQKAECLVLSA 1213
Qy 1581 -----QDRKLT-----STPRE-----IAKSP-----HST--- 1599
Db 1214 IDQVLGLADKTIIRGTETSEMRNFENNAEKVLIIPAIRMKVAEWQKQVAATPAAPVHATATP 1273
Qy 1600 --VPEHHPHDISP--VEHLLRGVSGVDLYRSHITPLAFDPTSIIPRGIPLDAAAAYILPRHL 1655
Db 1274 GALPANH-HTMAPATAPSAQRIATPESATORANPPGNSAVPRTITAAAPAP----- 1325
Qy 1656 APNPTVPHLYPYLIRGYPDTAALNRQTI-----INDVITSQOMHNTATAMAQADM 1709
Db 1326 TPAPLPLPRAAP--ISGHPAPSTVNTNNAAGARPVPAAPGPGVGHSGAIGVTAPV 1382
Qy 1710 LRG-----LSPRESSLALNYAAGPRGIIDLQVPH--LPV-----LVPP 1746
Db 1383 SRPPVTHGVVTAAPTTPR-SNLAASVPAG-----SMAQOHTSGAPAGSGNASISRAPP 1436
Qy 1747 TPQGTATAMDRLAYLPTAPQPFSSRHSSSPLSPGPHLTPTTSSSERERDRDRDR 1806
Db 1437 STAAPTA-----PTSVPVSVSTVRPMSYVTPGPGLTAPPTASSG-----ASAGY 1481
Qy 1807 DREREKSLTSTTTVEHA-----PIWRPGTEOSSSGSSGSGGGSSSRPASHSHA 1857
Db 1482 ARPANASTWPTISGQAMTSAVQSVSPRP--SSLTGGQGVPAIAAASRSPASGVYN 1540
Qy 1858 HOHSPISPRTDALQORPSSVLNHTGMKGIIITAVE-----PSKPTVLR--STSTSSVPRA 1910
Db 1541 PPASSLAPSTHSM--PSAVPTTASGAVSSTVSSLAATPLPSPAPRYGSPNNATPTPA 1597
Qy 1911 ATPPPATHCLPGTLDGVYFTLMEPVLKPKEARVARPERPRADTGHAFIAKAPARSGL 1970
Db 1598 TAIPRA---PL-----PAASSVSAPVTGQP-----SFGA--PASVPTP 1630
Qy 1971 PASPSKSGSP--RPLVPVSGHATARTPAKNLAPHASPDPPAPPASADPHREXTQS 2028
Db 1631 PTSASQAGPLSQPVGPAPAISSIGATPAASI-----PSSA----- 1668
Qy 2029 KPFSIQLELRSLGYHSGSYSPGCVPEVPSVSPSLTHDKGLPKHLEBKSHLEGELRP 2088
Db 1669 -----PAALAPVT----- 1676
Qy 2089 KQCPGVKLGGEAAHPLHLPLPE-----SOPSSSPLLQATPGVKGHQVVTLAQHI- 2139
Db 1677 -YPVQOASAAAAARLP-VTPAPAAHTTAQSVQVAPRPVTQSP-----VQSVADHVT 1726
Qy 2140 -SEVITQDTRHHHPQOLSAPLAPLYSPGASCPCVLDLRPPPSDLYLPFPDGHGAPARGSP 2198
Db 1727 SSQAST---TAHP--VAQSVPRVNSNPTISAPV-----AAVGTA 1762
Qy 2199 HSEGGKSPBPNKTSVLGGGEDGIEPVSPGEMTEPGHS--RSAYVLLYRDGQTPEPSR 2256
Db 1763 QVASAPVTQPAHRAHLLSSVSQSL-PQSVPHAAQAAAHQTPHSASREV-----PQSVPOS 1816
Qy 2257 MGSKSPGNTSOPPAFFSKLTESNSAMVSKKQEIKNKLNTNRNEPEVNTSQGTETFINM 2316
Db 1817 VPQATPOAVPRP-----STSLTPT-----AQOPGV---S 1843
Qy 2317 PAITGTGLMTYRSQAQVQEHASTNMGLEAIRKALMGKYDQWEEPPISANAFNPLNASAS 2376
Db 1844 PAVSGSGVPA--PSAAQSVAPAV-----SSTVPAAATVAPASTVAA 1883
Qy 2377 LPAAMPITAADGRSDHLYTSPGGGKAKVSGRPSRKAQSPAGLASCGRDPSPSVSHSE 2436
Db 1884 APTTRVTAAPALAS-AATNP-----APVPSQP-QHQITGOAP--AQOORPPAQAPPA- 1933
Qy 2437 GDCNRRPTLNTVRWEDRPSAGSTPFPVNPMLRLQAGVMASSPPGLPAGSGPLAGPHH 2496
Db 1934 -----TPTITSAAPRPPTLAPPPPPPP-----PPTEDPPPPPPPPPPPPPP- 1979

```
QY 2497 AWDEBPKPL 2506
Db 1980 -----PTPLM 1984

RESULT 67
Q63461
AC Q63461 PRELIMINARY; PRT; 2752 AA.
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Proline-rich protein precursor (Fragment).
GN PRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Spleen;
RX MEDLINE=89066721; PubMed=3198617;
RA Hemschoote K., Peeters B., Dirckx L., Claessens F., De Clercq N.,
RA Heyns W., Winderickx J., Bannwarth W., Rombauts W.;
RT "A single 12.5-kilobase androgen-regulated mRNA encoding multiple
RT proline-rich polypeptides in the ventral prostate of the rat.";
RL J. Biol. Chem. 263:19159-19165(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Spleen;
RX MEDLINE=92250652; PubMed=1577819;
RA De Clercq N., Hemschoote K., Devos A., Peeters B., Heyns W.,
RA Rombauts W.;
RT "The 4.4-kilodalton proline-rich polypeptides of the rat ventral
RT prostat are the proteolytic products of a 637-kilodalton protein
RT displaying highly repetitive sequences and encoded in a single exon.";
RL J. Biol. Chem. 267:9884-9894(1992).
DR EMBL; M86526; AAA41957.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT NON_TER 2752
FT SEQUENCE 2752 AA; 305057 MW; 55930CDE2818F7F8 CRC64;

Query Match
Best Local Similarity 3.3%; Score 438; DB 11; Length 2752;
Matches 564; Conservative 327; Mismatches 1071; Indels 968; Gaps 138;

QY 21 PHSLSYVQIARTHTDVGLEYQHSHRDYASHLSPGSIQ--PQRRPSPILLSEFQFGNER 78
Db 226 PHSYLP-----LQSQENAEAPHEFEQVELYQMETQTRNP-----EN 263
QY 79 SOELHLRPESHYLPDLKXSEMEFTSKRP-----RLELLPDLPLRPSPILLATGPAG--- 131
Db 264 SOQ-----EADYFPQSPDEDEPLIQEDPAHYQHHTLP-----TITGKPDVIDL 309
QY 132 ---SEDLTKORSLTKLEVPSPSP----- 153
Db 310 KITSPIKVSESLYEETPTQTPGPFVEAKLPSPSQOQPAETSPETPEGSSGIDLESSV 369
QY 154 -PHTDPELELVPPRLSKBELIOMDRVREITWQEQISKLKKQOQLE--EEAAKPPE- 209
Db 370 HPQEDSE-EIGLPTLQEDVDVSHLGPVLED-----ESSLESELPVLPESSEVSGSEN 424
QY 210 -PEKVPSP---PPIESKHSRLVQIIYDNRKKAERAHLEGLGPQVPLPN-OPSDTR 264
Db 425 QPEASVQPTVLPVLEQ-----EILFGAPGSSIET---VIE-----TLPIEIQPTQNE 469
QY 265 QYHENIKINQARKKLLIYFKERNARKQWKQFCQVDQLMLEALE--KKVERI--ENNP 320
Db 470 DYGQQLP-NVTVRPVDVALVTITSEPVKE--TESFLAPQEFVPHALEYNDVEFPVNEBEP 526
```

Db 1504 LELITTPQTFEGELS-QTVQES-----TTQNKEPKHEVAPFPVYQAVTPTFSQOAE 1557
Qy 1225 YRGSI-----THGTADLVLYKGTITR-----IIGEDSPSR-----LDRGR 1259
Db 1558 YQKSLQPLDLLELTVTSETKAYHSTISKNSLAINPQVHIQHPNAPBATVQPLDLLELTI 1617
Qy 1260 EDLS---PKGHVYIEGKGHVLSEBGMVSITQCKEDGRSSGPP--HETAAPKRTYDMM 1314
Db 1618 SSSLQPTAEGELLYSMQE-----TVTIQISEPPKQVVTVPVEQVAVPAPVQDQA 1667
Qy 1315 EGRVCRATSSASIE-----GLMGRA-----IPPER----- 1339
Db 1668 KYPLSSIVLSNLSLQDELTLSELLEGEAQLTTPDETVMVLPKDRQGIYPDHDHKKHLNLT 1727
Qy 1340 -----HSPHLKEQHIRGSITOGIPRSVVEAQED-----YLREAK-----LLKR 1380
Db 1728 EVTNQPFLEHTVQOPTIEERSQIOKTKTQITEPGKKVVPVLAQESSEVTIWPILKE 1787
Qy 1381 EGTPEPPPPRDLTAYKTOALGPLKJK-----PAHEGLVATVKEA--GRSIIHIEPREE 1432
Db 1788 TAFPTP-----HSMALQSLDEKLTIIHSHSPGWTQOHANLKEKSGHTTGKILLDY 1836
Qy 1433 LRHTPELPLAPPLKEGSIQTGTPLYDTGAST-----TCSKGHDVRSLLIGSPGR 1482
Db 1837 AEPNMEIELKHGHLFLKTTTEATTESNTQMTKSLKQVTFALTQNKKSMLPALVESQDE 1896
Qy 1483 TPPP-----VHPLDVNADARALERACYEESLKSRRPGTASSSGGSIARGAPVIVPELGRQR 1538
Db 1897 SQPPNMSLQPLD-----QELTSSQP----- 1918
Qy 1539 SPLTYEDHG-AFFAGHLPRGSPVTWREPT--PRLOEGSLSSKASQDKLSTPREIAKS 1595
Db 1919 -----HGWPVHIENTEKILYHVAEPPTGPFVEPELFLTKITKSRPVQGTATQMAAS 1971
Qy 1596 PH---STVPEHHHPHISPYEHLRGVSGVDLYRSHIPLAFDPTSIPRIGIPIDAAAAYLP 1652
Db 1972 PREMVSRAPEK-----EAVLSG-PGEDQDESPPNMSLQSLQDLQELTLSSQPHGWIP 2023
Qy 1653 R-----HLAPNTPYHLYPYLI-----RGYPD-----TAALENROTIINDYITSQ 1693
Db 2024 HPNTHGKIYHVAEPPTGPFVEPDPLFLTKITKSPVEWTLTRTDKSKEMV-----SQ 2078
Qy 1694 QMHNHTATAMAQRADMLRGLSPRESSL-----ALNYAAGPRGIIDLQVPHLPVLVPPTP 1748
Db 2079 SPKYEEAVLPVHGEQESRSPPNMSLQSLQELTLSSQPHG-----WVPH-----PPNTH 2129
Qy 1749 GTPATAMDRLAYLPTAPOPFSRRHSSSPLSGGP-----THLTKTITTTSSSRE 1797
Db 2130 GK-----IYL-----HYAEP--PTGPFVEPDPLFLTKITKSPVQGTATRMV 2169
Qy 1798 RDRDRDRDREREKSLTSTTTTVEHAPVIRPQTQSSGSGSGGGGSSRRPASH--- 1854
Db 2170 KSPKEMVSLDPENKEAVF-----PAQEGKGESFPSS---PNMSLQSLDHLF 2213
Qy 1855 --SHAHQHSPISPRTOALQORPSVLH-----NTGMKGIITAVEPSKPTVLTRSTSTSPVR 1908
Db 2214 MSSQPHGWLPHPPKTPDKI-----YLHVAEPPTG-----PVEPDLFLFATTK-SKPVQ 2262
Qy 1909 PAATPPATHCPGLGTGCVPTLMPEVLLPKEAPRVARPERPRADTGHAFKAPPARSG 1968
Db 2263 GTTT-----EMAKSPKEMVS-QTPEYK-----EAVLSGP-----G 2291
Qy 1969 LEPASSPSKGSFRLPVPVSGHATIAARTPAKNLAPHASDPDP-----APPA 2016
Db 2292 EDQDESPS-----PNTSLKSLDQEVAMSSQPHSGVPHPPKTPGKIYHLSIEPPPG 2342
Qy 2017 SASDPH-----REKTQSKPFSIQLELSRLGVHSGSYSGVEGVEPVSPVSSPLTHDKGLP 2071
Db 2343 PFKVPTDLILVKTITKSPAETWTPRRIDKLLKEMVPHSPVEEAV-----FP 2389
Qy 2072 KHLEBLDKSHLEGERLPKQPGVPLGGEAAHLPHLRPLPESQPSSSLQTPAGVKGHQR 2131

Db 2390 AHGEGQDES-----GSPNMP-LQPLDQELTLSSQ----- 2418
Qy 2132 VVTLAQHISEVITODYTRHHHPQQLSAPLAPLAPLSPFPGASCPLDLRLRPPPSDLYLPPPD-- 2189
Db 2419 -----PHGWVPHFN-----TPGKIY-----LHYAEPPTGPFVEPDLF 2452
Qy 2190 -----HGAPARGSPH-----SEGKRSPEPNKTSVLGGEDGIEPVSPPEGMTEPGHSR 2238
Db 2453 FLRTTKSPVQSGPKETAKSPKEMVSTPYKEAVLSGPGEDQDESPPN----- 2503
Qy 2239 SAVVPLRYRGEOTEPEPSMGS-----KSDG-----NTSOPPAFFSKLTE--SNSAMVKS 2285
Db 2504 ---MSLSLQDOEVTMSQPHSGVPHPKTPGKIYHLSIEPPPGFVFKPTDILILVKTITKS 2560
Qy 2286 KKQE-----INKLNTNHRNEPEYNIQPGTEIF-----NMPAITGTGLMT 2326
Db 2561 KPAETWTPRRIDKLLKEMVPHSPYE-----EAVPHAEGQDESGSPNPLQPLDQELT 2615
Qy 2327 YRSQA-----VOEHASTNNGLEAIRKALMGKYQDWEESPPLSANAFNPLNASLUPAAMPI 2383
Db 2616 LSSQPHGWVPHFNPT-----PGKIYHVAEPPTGPFVEPDLFLTRTPKSKPV 2663
Qy 2384 TAADGRDHTLITSPGGGKAKVSGRPSRKAAPGLASGD--RPSVS 2431
Db 2664 ---QGTPTQMAKSP---EEMVSLSPKNKETVFPQAGKQDESPPNIS 2706

RESULT 68
O57580 PRELIMINARY; PRT; 1151 AA.
ID O57580;
AC O57580;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE High molecular mass nuclear antigen (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98034404; PubMed=9365273;
RA Shimada K., Harata M., Mizuno S.;
RT "A nuclear matrix-associated high molecular mass nuclear antigen,
RT HMNA, of chicken and marked decrease of its immunoreactivity during
RT the progression of S phase."
RL J. Cell Sci. 110:3031-3041(1997).
DR EMBL; D88440; BAA24137.1; -.
DR PIR; T18535; T18535.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1151 AA; 109708 MW; 2403F6835F9A2AB3 CRC64;

Query Match 3.3%; Score 436; DB 13; Length 1151;
Best Local Similarity 19.7%; Pred. No. 8.6e-13;
Matches 346; Conservative 154; Mismatches 548; Indels 706; Gaps 77;
Qy 752 PSPTHEAKUTGQNGKPP---ATLGADGPPG-PPTPRR-TSRPIETPPASEATGAP 806
Db 31 PRPKVPIAELHPAOPPPKVPVPIGGAPPPGTEPTPSKPTDGADAAKASAEIT--- 87
Qy 807 TTPPAPSPSAPPVVPKEKEETAAAPVVEEGEQKPAAEELA----- 852
Db 88 SPPPASFP-PDGPKAPSGAGEAE-AGTPPSPQFAGTTPPSQAGAPKGDGTAQPSGT 145
Qy 853 ---VDTGKAEEPVKSECTEEAEAGPAKQDAEAAEATAE-----GALKAKKEGSGSRAT 904
Db 146 KSGADGKPAQDVKPAATTAATEARPAASPTVPKATAEATAVTAASQSPKAAATDAAV 205
Qy 905 TAKSSGAPQSDSATCSADEVDEAGGDKNRLSPPRSLTPTGDPANA-----S 956
Db 206 TAASQSPKAPKATVEVKPAAAVAKEK-----AVTAAAAAKATAEAKPAPVTS 253

QY 957 POKPLDLKQKORAAAP-PIQVTKVHEPPREDAAPTTPAPP 1000
Db 254 PTIPCSSAEAKPLTAASPTASKATAEAKVPVATASLMATKV-----TAEAKPA-PSFS 305
QY 1001 PPNQLOPESDAPQOGSSPRGKSPAPADKEAFAAEAQKLPDGPWCWTSGLPPVPDR 1060
Db 306 VP-----KATYTDKAVTATAPK-----AGPDVKPAVAVCAEAKPAPP-----PPQ 346
QY 1061 EVIKASHPAPDPSAFSYAP-----PGHPLPLGLHDTARPVLPPTISNPPPLISSAKHPSV 1117
Db 347 QLPKAAA-AAAPTGTTELKAPATPPHGPS-----RANSHTVTVTPNV-----PRAAATVETA 398
QY 1118 LERQTGAISQGMVOLHVYPYSHAKAPVGPVTMGLPMDPKLAPFGVQEQSLSPRGQ 1177
Db 399 -----GAVPKAST-----GTTFAAAPOQPV-----PKAAPVTPPPPOQAVPR-- 435
QY 1178 AGPPESLGVPTAQEASVLRTALGVSFGSITKGPSTR-VPSDSAIYRGSIHTGTAD 1236
Db 436 -----ATAAAAPV-----TPQOPVTKAATTNATPPPOPIPKAATTTTATFV- 478
QY 1237 VLYKGTITRIIGEDSPSLDRGREDSLPKGHVYBGGKHVLSYEGGMSVTQCKEDGRS 1296
Db 479 -----TP-----QOPIPKA-----GTD 490
QY 1297 SSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAPPERHSPHLKEQHIRGSIT 1356
Db 491 AAPPB--AVPKAPSD-----GRAATP----- 509
QY 1357 QGIPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTQMLKPKAHEGLVA 1416
Db 510 -GVNNAATDQK-----PPPTQS----- 527
QY 1417 TVKEAGRSIHETPRELRHTPELPLAPRLKEGSIQTGTP---LKYDTGASTTGSKKHDV 1473
Db 528 -----VPSAVTEPKQPRAAPPPSNEATPAVPSFNLK----- 561
QY 1474 RSLIGSPORTPPVHPHLDVMDARALACRYEESLKRPGTASSGGSIARGAPVIVPEL 1533
Db 562 -----SPLPTIP--KPVPLMA-----LTPQPVTAQ-----MVTQL 589
QY 1534 GKPSQSLTYEDHGAPFAGHLPGRGSPVTRGPTPRLQBGSLSSSKASQDRKLTSTPREIA 1593
Db 590 AATKPSPT-----VPKASFKAALMTPPP-PPGLPRALAAKLLGLPSSP--VA 634
QY 1594 KSPHSTVEHHPHPTSPYEHLLRGVGVLYRSHIPLAFDPTSPRGIPLDAAAAYLPR 1653
Db 635 SAMHAKVTP-RPLPASP-----VPMASPASLGP-----DAARV----- 667
QY 1654 HLAHPNTYPPHYLPYLRGYPDTALENROTIINDYITSQOMHNTATAMAQRADMLRGL 1713
Db 668 ALATNAASPGAKP-----EAGGNGTLMAPMG----- 694
QY 1714 SPRESSALNVAAGPRGIDISQVPHLVL-----VPTTGTATAMDRLAYLPTAPQPSS 1770
Db 695 -----AANTQMAPIGAAGAAQTAPMGAAHTHSPMGAGGATQW-----SPTGAAN 739
QY 1771 RHSSPLSPGPTHTLTKPTTTSSSERDRDRDREREKSIILSTTTTVEHAPIWRPG 1830
Db 740 TH-MSPIGAGGATQMS-PMGAAANTQMS-----MGATTTQMSPMGAAA 780
QY 1831 TEQSSGSSGGGGSSRRPASHAHQHSPIPRTDALQQRPSVLHNTMGKIITAV 1890
Db 781 TTQSPSM-----GAAATQVATTSAGNTQVSPMGAAATP---PQTPSV-----GAATTP 825
QY 1891 EPSKPTVLRSTSSPVRAATFPFATHCPGLGTLGDVYPTLMEPVLLPKAPRVARPER 1950
Db 826 QPS--PMGAATLMSPMGAATTPQS-----PMGAV-----TTQPP-----PMAATNTTQPP 871
QY 1951 PRADT-----CHAFAPAPARSGLPASPSSKSGSEPRPLVPVSGHATIAATPAKNLAP 2004
Db 872 MAASTPQSTPMGAATTTQSPPMGATTTQSPPMGASTPO--APP-----TVAGSP----- 918

QY 2005 HHASPDPPAPPASADPHREKTQSKPFSIOBLELRSLGYHGSSYSPGCVPSVSPSSPL 2064
Db 919 ---TPPPPIPPSPTA-----QTSPOPM-----KSPPPDP-- 945
QY 2065 THDKGLPKHLEELDKSHLEGELRPKQPGVKLGGEAAHLPHLRPLPESQPSPLQATAP 2124
Db 946 -----PKAFSAAAOQTSPAHHAN-----ASP 966
QY 2125 GVKGHQRVVTUAQHISEVITQDTRHHHPQQLSAPLPAPLYSPFGASCPCVLDLRRPPSDLY 2184
Db 967 GV-----TAVSPAP----- 976
QY 2185 LPPPDHGAPARGSPHSGKSPENKTSVLGGEDGIEPVSPPEG-----MTEPGHRSAY 2241
Db 977 -----GVTEASPSADGARGLSPGTAAT-----DG--PKASPAATADVTEAATDVTA 1021
QY 2242 YPLLRYDGEQTEPERMGSKSPGNTSQPPAPFSLKTESAMVKSQKQINKLINTHNE 2301
Db 1022 ATAVPAEAAPTAKRSSSSSSSSSS-----SSSSSSSSSSSSSSSSSSSSSPAS 1075
QY 2302 PEYNISQPGTEIFNMPAITGTGLMYRSQAVQEHASTNMGLEAIRKALMGKYDQWESP 2361
Db 1076 P-----APAVG-----DQOQMT 1088
QY 2362 PLSANAFNLPLNASLASLPAAMPITAADGRSDHTLTPSGGGKAKYSGRPSRKAAPSAPGL 2421
Db 1089 PGAAQSVPPVTEAAVQEAASAAAAAG-----AEREGRPTRRKRTRSSSS 1134
QY 2422 ASGRDPPSVSVHS 2435
Db 1135 SSSSSSSSSSSSSSS 1148
RESULT 69
Q81Q87 PRELIMINARY; PRT; 9196 AA.
AC Q81Q87;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CG32377-PA.
GN CG32377.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova K., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Bantz J., An H., Baldwin D., Banson J., Beeson K.Y., Bueson D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragov V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003558; AAN12045.1;
DR FlyBase; FBgn0052377; CG32377.
SQ SEQUENCE 9196 AA; 1008711 MW; E253273B32DBBE88 CRC64;
Query Match 3.3%; Score 434.5; DB 5; Length 9196;
Best Local Similarity 19.0%; Pred. No. 1.4e-11;
Matches 537; Conservative 386; Mismatches 1083; Indels 819; Gaps 130;
QY 55 FGSIIQQRRLPSLLSEFPQNGERSQELHLPESHSLP-----ELCKSEMEF 102
DB 3562 PKPLDKPIHSPTSLVTSVTGDKSS---LHPEKPKSVKTEKVLAKPDDISKS---V 3615
QY 103 IESKEP-----RLELLDPLLRP---SPLLATGQPAQSEDLTKDRSLTGKLEP 147
DB 3616 VETDKSPKEYSDDETEDEIE-IPKLDKPIHSPTSLVTSVTGFG-----VDKS-----P 3663
QY 148 VSPSPPHPTDELELVPPRLSKEELIQNWDRVDREITVVEQIQIKLKKQQLBEEAAKP 207
DB 3664 LNPEKPN-----PEKDKVLAKPDDISKSVKTKPIPK-EVSDDETDDEIEIP 3714
QY 208 PEKPKVSPPP-----PI---ESKHSILVQIYIDENRKKAFAAHRILEGLGPQ 251
DB 4637 D-----ENDDEIDFPKLDKPIHSPTSL--VTGVTGDKSSSLHP---EEXPKSPKXDEK 4687

Db 3715 KPLDKPIHSPTSLVTSVTGATFGDGSPLHPEKPKS-----PEKDKVLAKPDGSSKS 3766
QY 252 V---ELPLYNQPSDTRQYHENIKINQAMRKL-----ILYFKRNHARKQWKQKFC 299
DB 3767 VVETDKSPKEYSD-STEDEIEIPKLDKPIHSPTSLVTSVTGDMKSPHLPBEKNSP 3825
QY 300 QRYDQLEALEKVKVERIENPRRRAKESK--VREYEEKQFPFIRKQRELMOSRQVQR 357
DB 3826 EKXD-----EKVLAKPDDSSKVVVTKDKPIKEYSD---DETEDEIEIPKLDKPIHSP 3876
QY 358 GSGLSMSAAKSEHEVSIIDGLSEQNLEKQMLAVIPMLYDADQOIRKFINNGLMA 417
DB 3877 TS-LVTSVTGDKSS-----LHPEKPKSPEKDKVLKPKDDSSKSVVK----- 3921
QY 418 DPMKVKYKDRQVMNMWMSQEKETFRE-----KFMQHPKN-----FGLASFL---BRKT 462
DB 3922 -----TDKPIKEYSDDETEDEIEIPKLDKPIHSPTSLVTSVTGVDKSPMLNPEKPN 3975
QY 463 VAECVLYYYLTKNENYKSLVRSYRRRKSQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 520
DB 3976 SPEKDKVLAKPDDISKSVVK-----TDKPIKEYSDDETD 4012
QY 521 DE-----KEKEKAEKEE-----KPE-----V 538
DB 4013 DEIBIPKLDKPICOPTSLVTSVTGATFGDGSPLHPEKPKSPEKDKVLAKPDGSSKSV 4072
QY 539 ENDK-----EDLLKEKTDGSCDNDEKAEVASKGRKTANSQGRKRGRITRSMANEAN 591
DB 4073 ETDKSPKEYSDDETEDEIDFPKPKPFKEATPS-----VT 4109
QY 592 SEEAITPQQAELASMEINLESSRWTEEMETAKGL-----LEHGRNWSAIA 638
DB 4110 PVSTIPDKALDFVSKAECSITTIQNIKLASITTKDKIVQPIDSLSNLKD-DKFP 4168
QY 639 RMVSGKTVSQCKNFYFNKRNQNDLEILOQHKLK-----EKERNARKKKKAPAAA 690
DB 4169 TSVGDKAKSPRESVKPNLKEYSKKEKPDSPHPSLVTSVMGSGDKSPHLPBEKPKSPEKK 4228
QY 691 SEEAAPP-----VVEDEMEASGVSGNR--EEMVEAEALH-----ASGN 729
DB 4229 DEKVLKPDSSKSVVETDKSPKEYSDDETDDEIDIPKALDKPIHSPVSLVTSVMGSGD 4288
QY 730 E-----VPRGECGSPATVNNSSDTEIPSPHTEAAKDTQNGPKP- 769
DB 4289 KSPLHPEKPKSPEKDKVLKPDSSKSVVETD---KPIKEYSDDETDDEIEIPKPL 4345
QY 770 -----PATLGD-----GPPGPTPTPRRTSRAPIEPTTPASEATGAPTPPAPP 813
DB 4346 DKPIHSPTSLVTSVTGVDKSPHPEKPKSPEKDKVLAKPDDSSKSVVTKDKP---- 4401
QY 814 SPSAPPVVPKKEKEETAAAPVVEGEEOKP---PAAELAVDTGKAEPEVKSEC-TEE 869
DB 4402 -----IPKEYSDDEND-----DEIDFPKLDKPIHSPVSLVTSVMGSGDKSPHLPBE 4448
QY 870 AEGFAKGDAAEAAEATAEGALKAEKKEGSGRATTAKSSGAPQD--SDSSATCSADEVDE 928
DB 4449 KPKSPEK-KDE-----KVLKPDSSKSVVETDKPIKEYSDDE---NDDEIDF 4493
QY 929 AEGGDKNRLISPRSLTPT---GDPRANASPKPLDLKQLKQRAAIPPIQVTKVHEPP 985
DB 4494 PKALDKP--ISHTSLVTSVTGDKSSSLHPPEKPKSPEKDKVL-----PK 4539
QY 986 REDAAPTKPAPPAPPQNLPESDAP-----QQFGSSPR-----GKSRSPAPPA 1030
DB 4540 PDDSSKSVVETDKSPKEYSDDETDDEIEIPKLDKPIHSPTSLVTSVTGVDKSPHLPBE 4599
QY 1031 DKEAFAAEAKLPGDPCWTSGLFPFVPPREVIKASPHADPPSAFSAFPPHPLPLGLHD 1090
DB 4600 EK-----PKSPEKDKVLAKPDDSSKSVVTKDKPIKEYSD 4636
QY 1091 TARVLPRPPTISNPPPLISSAKHPSVLERQIGAISQMSVOLHVPVSEHAKAPVGVPTM 1150
DB 4637 D-----ENDDEIDFPKLDKPIHSPTSL--VTGVTGDKSSSLHP---EEXPKSPKXDEK 4687

Db 887 --PSLPPTQ-----EVEEEGPMQWSPNQVPASH-QTSPTASRGHAMPHSRRTTTPPOS 939
QY 1387 --PPSRDLTEAYKTQALQPLKKAHEGLVATVKEAGRSIHEIPREELRHTPELPPLAPR 1444
Db 940 TVQPTVSPQALYSVQ---PAASASHE-----FAPPPRSQTSQALYGNRVKPIEPI 991
QY 1445 PLKEGSIQTGTLKDYTGASTTGGSKKHVRSLLIG-SPGRTFPPVPHPLDVMADARALERAC 1503
Db 992 P-RPSSVTDAT-----SPRAALYSPIWTQPSSTFPI-----1022
QY 1504 YEESLKSRPCTASSGGSTA-----RGAPVIVPELGRQSPPLTYEDHGA--1548
Db 1023 --STTSRPRGNTQNFNLVPPDTGREHDPLOQRWGVPLISWGV-----GALV 1068
QY 1549 -PFAGHLPR-----GSPVTMRBTPRLQEGSLSSKASQDKLTSTPREIAKSP-----1596
Db 1069 TFPKXNVRYGMQOTAPMVRSP-----GEVVKSMKDIEPMEERLAKFPGLKKG 1119
QY 1597 -----HSTVPEHHPHPIPSYEHLLRGVSGVDLYRSHIPFLAFDPTSIPR 1639
Db 1120 SKKETIAWLTNGIEMLETRVPHNLQNPSHDDKRTTERLLMK--ILRVF-----1170
QY 1640 GIPLDAAAYLPRHLAPNPTYPH-----LYP-PYLIRGYDPTALENQTIIINDYITSQ 1693
Db 1171 -VEHDGV-----LEGNPTVQAVREILYPEASTIGQCPFANALNPSGMSGNATSLS 1221
QY 1694 QMHHTATAMAO-RADMLRGLSPRESSLALNVAAGPRGLIDLSQVPHLPVLVPPPTGTPA 1752
Db 1222 QADSVDSAVEGIRNHLISG-----DNEKAV-WAA-----VDKRLMGHAFLL-----A 1263
QY 1753 TAMDRLAYLTPAQPF-----SSRHSSPSLS-----PGG-----PTH-----LTK 1787
Db 1264 NALNPDLYKRYA-QEFVKNEVNSTGHNESLAALYDVLSGNHEESVDELVPAHARAGLQ 1322
QY 1788 PTTTSSSRERDRDRDREREKSLTSTTTVEHAPIWRPGTEQSS-----1835
Db 1323 VAKNSSGSKDAMGLDKWRETLILLNRTADDARAINFGLNLLSGYGRAEAAHICFL 1382
QY 1836 -GSSGSSGG-----GSSSRSPASH-----SHAHQS-----PJS-1864
Db 1383 FARSTQIFGLDNPNNSFVLVGSDDRHOADHPFAKIEPLLSSEVYEQSLAGGTVPSVN 1442
QY 1865 PRTOALQORPSVLHNTKMGK-IIITAVEPSKFTVLRLSTSTSPVRPAATFPATHCPGLG 1923
Db 1443 PHLAAYKLQHAVALAEYGFCDKALQYCEAITAITAQTKRS-----PYHPILEA 1492
QY 1924 TLDGYVPTLMEVLLPKZAPR--VARPERPRADTG-----HAFKAPPARSGLEPASSPS 1976
Db 1493 YVDDLMLKRLQ---APKEESNWIKPSNMKVSDSMNRFNKVFAGDDNEDGSK--GSPD 1547
QY 1977 KGSEPRPLVPVSGHATARTPAKN-----LAPHASPPDPAP--PASASDPHREKTQ 2027
Db 1548 AAGESGPFARIAGTGTTSRSPSVNLTFTGATIPSYGMPGAPVTNGFMFSPPPPTRTA 1607
QY 2028 SKPFSIQLELRSLGYHSGSSYSPGVPVSPSSPLTHDKLPK-HLE-----ELDKSHL 2082
Db 1608 SR-----YAPGAPQSTPNYNPYETNSYPAPSSMERASGEYSRSV 1649
QY 2083 EGELRPKQPGPKVG-GBAAHLPHLRPLPES-----QPSSSPLLOTAPGVKGHORV 2132
Db 1650 E---LPRQSLDSQRGYSHSSVAPNRTSSPAQPTPYGTPQESSYSLHMOP-----QQSL 1702
QY 2133 VTLAQHISEVITQDTRHHQ-OLSA-----PLPAPLYFPFGACFPVLDLRPPSDLY 2184
Db 1703 TSPA-----ATTSGYQPTPQNNNSANDENEPESAPSTGY-----QPSSYGY 1746
QY 2185 LPP--PDHCAPARGSPHSEGGKRSPEPNKTSVLGGGEGDIEPVSPPEGMT---EPGHSRS 2239
Db 1747 EPPSTPYEATNDEEBGASKENDSNQ-----GGE-GVNTFEPSPQPSYEPSPPEP 1800
QY 2240 AVYPLLRYDGEQTEPSRMGSKSPG-----NTSQPPAFPSK--LTSNSAMVKSCK 2287

Db 1801 DTPP--SKDDQDSEBEKPKKKGKPMYDDDDDDFFAAPKPAKSKAEIDRENEEMVRRIA 1858
QY 2288 OEINKKLANTHRNEPEYNIISQPGTEIFNMPAITGTGLMTYRSQAVQEHASTWNG-----2341
Db 1859 EEEAKRAAEAK-----AKKGMGFTSWFAKKEAAAADANAAGSSPOK 1900
QY 2342 -LEAIRKALMGKYDQWEEESPPLSANAFNPNASASLPAAMPITAADGRSDHLTSP---2397
Db 1901 PIRAKLGEANSFYDYP-EQKEWINKNA-SPEDQAACKSTPPPKGKIPRSSASSPAPMG 1958
QY 2398 ---GGGKAKVSGRPSRKAASPAKGL-----AS-----GDRPPSVSVHSEGCNRRTPUT 2446
Db 1959 MGVGGSAPNTPGRASAPPTGPRPAALMPASSENVSNGSPSAVGLPSGSGN-----2012
QY 2447 NRWVEDRPSASGTPFPYNPLIMRLQAGVMA5PPPPGLPAGSGPLAGPHAWD-----2499
Db 2013 -----GPPSAGLLSPGMGAAMORPASTSTSGPPAGT---SKPLSATSSIDDLIGAAMP 2063
QY 2500 ---EEPKPLLC5QY 2510
Db 2064 RKGEEKPKRAARY 2078
RESULT 73
Q9BX49
ID Q9BX49 PRELIMINARY; PRT; 1404 AA.
AC Q9BX49;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE BG174L6.2 (MSF: megakaryocyte stimulating factor).
GN BG174L6.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133553; CAC36090.1; -.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;
Query Match 3.2%; Score 429; DB 4; Length 1404;
Best Local Similarity 20.8%; Pred. No. 2.4e-12;
Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;
QY 476 NENYKSLVRRSYRRRGKSKQ-----QQQQQQQQQQQQQQQPMRPSQOEKDEKE 524
Db 126 SQTIKSTTKRSPFNKKTKKVIIESEITEHVSQSSSSSSSSSSSSSSSSSSSSSSSSSSSS 185
QY 525 KEKEAEKE-EKKEPVENDKEDLLKEK-----TDDT--SGEDNDEKEAVASKGRKTANSQ 575
Db 186 KNSAANRELQKLVKONKNRTKKKTPKPPVDEAGSLDNGDFKVTTPDTSTQHNK 245
QY 576 GRRKGRITRSMANSEEAITP-QQSAELASMEINSSRWTEEMETAKKGLLEHGRNW 634
Db 246 VSTSPKIT--TAKPINRPSLPPNSDTSKTSLATVNKETTETKETTITNKTSTDGKEK 303
QY 635 SAIRAMVGSKTVSCKKPFYFNKRNQLDELLOQHLKWE---KERNARRKKKKAAPAAAS 691
Db 304 TTSAKETQSIKTSKOL-----APTSKVLAKPTKATTTTGPALTTPKPTPTTPK 356
QY 692 EEAAPPVVEDEMEASGVSQNEEMVEEAEALHASGVNEVPREGCSGPATVNNSSDTESI 751

Qy	920	T-CSADEVBAEGDKNRLLSPRPS-----LTPTCGDPNANASPOKPLDLKQLKORAAAIPP	975
Db	699	TQCFL-----ENQGLEAVQSMEDPTQAFMLTP-----POE-----LGP	732
Qy	976	-----IQVTKVHEPPREDAATPKAPPAPPQNQLQPEDAPQOQSPSGKRSRSPAPPAD	1031
Db	733	SHCSFQTTGLDPEWVLA-----TQFCURESEDETQFDT-----H	771
Qy	1032	KEAFABEAQKLPDGPWCWTSGLPPVPPPREVIKASPHAPDPSAFSAPPGLPLGLLHDT	1091
Db	772	LEAYG-----PCLS-----AIPGDQHPES-----PVHTPEMGLQCR	805
Qy	1092	ARVLPRPPTISNPPPLISSAKHPSVLEROIGAISQGMVOLHVHPVSEHAKAPVGPVTMG	1151
Db	806	GR-----QTVDKVMGIP-----KETAERVGP-----ERG	829
Qy	1152	LPLPMDPKKLAP-----FSGV-----KQEQLSPR-----CQAGPPE-----	1182
Db	830	PLERETEKLLPERQTDVTGEEELTKGQDREKQQLLARDTQRESKNGESASPERDRE	898
Qy	1183	--SLGVPTAOEASVLRGTALGSPVGGSIITKGIPTSRVPSDAITYRGSITGHTPADLVYK	1240
Db	889	SLKVEIETSEIQ-----EKQVKQTLPSKA-----FE	916
Qy	1241	GTTRIIG--EDSPRLDRGREDSLPKGVHVIYEGKKGHVLSYEGGM-----SVTQCSKED	1293
Db	917	REVERPVANRECDPAEL-----EEKVPKVLERDQTQGEF-----EGGSQDKQGOASSPTPEP	969
Qy	1294	GRSSGPPHTAAPKRYIDMMEGEVGRAIISASIEGLMGRAIIPERHSPHLKEQHHRG	1353
Db	970	GVGAGDLPGTTSAP-----VPSQSGGRGSPVSPRRHQ-----KG	1005
Qy	1354	SITQIGP---RSYVEAQEDYLREAKLLKREG---TPP-----PPPSRDLTEAYKTQ	1400
Db	1006	LINCKMPPABKASIRAEKVSRGD-----QESPDACLPTVPEAPAPKQPLNSQSQKH	1060
Qy	1401	ALGPLKPKAHEGLVATVKBAGRSIHIPREELRHTPELPLA-----PRP-LKEGSIIT	1452
Db	1061	LAPPELLSP---LLPSIKPTVRKTRQDQSQE---APEAPLSSELEFPFKPKIRTRKSS	1113
Qy	1453	QGITPLKYDTGASITGSKKHVRSILGSPGRTFPVHPPLDVADARALERACYESLSKRP	1512
Db	1114	RMTPPP---ATSAAPEHPFSTS-----TAQEVTPKPT-----SQATRSR-	1149
Qy	1513	GTASSGSGSIARGAPVT--VPGLKGRQSPPLTYEDHCAFGHLPGRGSPVTMEPTPRLQ	1570
Db	1150	-----TNRSSVKTPEVVPTAPEL-----QSTSTD-----OPVT-SEPTSQVT	1187
Qy	1571	EGSLSS-----KASQDRKLSTSPREIA---KSPHSTVPEHHPHIPSP	1610
Db	1188	RGRKRSRSVKTPETVPTAPLELOPSTSTDPTVPTSEPTSQATGRKNRSSVKT--PEPVVP	1245
Qy	1611	YELLRGVSGVDLYRSHIPLAFDPT-----SIPRGIPLDAAAAAYLP-----RH	1654
Db	1246	TAPELOPSTSDQ-----PVTSEPTYQATGRKNRSSVKTPEPVPTAPELRPFSTSDRP	1300
Qy	1655	LAPNPTVPHLYPPVILRGYPDTALENNQTIINDYITSQOMHNTATAMAQADMLRGLS	1714
Db	1301	VTKPPT-----SRTRSRTNMSSVKTPETVVP--TAPELOJSTSTD-----OPVT	1343
Qy	1715	PRESSLALNVAAGPRGIDLSQVPHLPVLVPPTPG-TPATAMDRLAYLPTAQOPF----	1768
Db	1344	PKPTSRT-----TRSRTNMSSVKNPSTVPIAPELPPSTSTEQ-----PVTPEPTSRA	1393
Qy	1769	-----SSRHSSSPPLSPCGPHLTKPTTSSSERDRDRDRDREREKSIILSTTIVEH	1823
Db	1394	GRKNRSSGKTPETVLVPTAPK--LEPSTSDQVPTPEPTSQATGRNRSVKTPEVVPT	1451
Qy	1824	APIWRP--GTEQSSGSGSGSGGGGSSSR-----PASHASHAQHPGI-----S	1864
Db	1452	APELOPSTSDQVPTPEPTSQATGRTRDRSRSVKTPETVPTAPELQASSTDQVTSDDPT	1511
Qy	1865	PRIQDALQORPSVLHNTGMKGIIITAVBPSKPTVLRSTSTSPVRPAATFPFATHCPGLG--	1922

Db 58 RMPDCSVALPPFSISKQHAEBIEIILAWDKAPILRDCGS-LNGTQILRP-----PKVLSPGVS 112
Qy 75 GNERSQEL-----HLRPESHV---LPELGKSEMEFIESKPRLELLPDPILLRSPILLATG 127
Db 113 HRLRQDELILFADLLCQVHRLDVLSPFVSRGFLTVETPRVQGETQORLL-----163
Qy 128 QPASSEDITKDRSLTGKLEPVSPSPHPTDPELELVPRLSKBELIQNMQRVDRDREIMWE 187
Db 164 -----LAED-----SEEVFLSER-----178
Qy 188 QOIKLKKKQOOLEEAAKPEPEKPVSPPPPIESKHSRLVQI1YDENRKAEAHRILEG 247
Db 179 -----RWKKSRTSSSVIPESDECHSP-----VLGG 207
Qy 248 LGQVELPLYNOPSTROYHENIKINOAMRKLILYFKRNHARKQWKQFCQBYDOLME 307
Db 208 LGPPE-----AFNLNSDT-----DVBE 224
Qy 308 ALEKVERIENPR-----RAKESKVREYKQFPEIRKQRELQERMQRVGOGRSG-----360
Db 225 GOQPATEAASAARGATVEAKQSEAKVVTIQL-----KQPLVKERDNDTKVKGAGNGV 282
Qy 361 -----LMSAARSEHEVSEIID-----GLSEQENLEKQMLQAVIPMLY-----DADQ 404
Db 283 VPAGVILERSQPPGSDSDTDVDDSRPPGRPAEVHLERAQ-----PFGFIDSDTDAEE 335
Qy 405 ORIKINNGLMADPMKV-YKDRQVNN-----MWSEQKEITPREKFMQHPKNFG 452
Db 336 ERIP-----ATPVVPMKRIKIFHGVGTRGPAAGLAHLOSQAGSDTDVEEGKAPQ 387
Qy 453 LIASFLEKRTVAECVLYVLTAKNENKYSLVRRSRRRGKQQQOQOQOQOQOQOQOQPM 512
Db 388 AVP-----LE-KSQASWINSDTDDEEVSAAALTLAHLK-----ESQPAIWNDAEDM 435
Qy 513 PRS-----SQEKEDEKEKEAEEKEKEPVENDKEDLLK-----EKTDTSGE 556
Db 436 PORVLLQORSQTTTERDSDTDEEBELPVEN-REAVLKDHKTIRALVRAHSEKQDPPFGD 494
Qy 557 DNDEKEA-VASGKRTANSQGRKGRITRESMANEANSSEAI-----TPQQS-- 601
Db 495 SDDSVEADKSSPGIHLERSQASTVDINTQVEKEVPPGSAI1HIKKHOVSVEGTNQTDKV 554
Qy 602 -----AELASMEINLESSRW-----TBEEMETA-----KKGLL-----EHGRNWSA 636
Db 555 AVGGPAKLLVSLAEA-WPLHGDCECTDAEGTALTASVADVRSQSLPAEGDAGAEEWA 612
Qy 637 I-----ARMVGSK---TVSQCKNFYNYKKQNLDEILOQHLKMKERNARRKKKAP 687
Db 613 AVLKQERAEVGAQGPPVAVQVEQDL--PISRENLTDLVVDTDLTGSESTQPOR-----663
Qy 688 AAASEBAFPVVEDEMEASGV-----GNEEMVEEA-----EALHASGNEV 731
Db 664 -----EGAQVPTGREHQVGTGKSDNYGSEDLDLQATOCFLENQGLEAVQSMEDEP 718
Qy 732 PR-----GECS-----GPATVNNSSDTESTIP-SHPTAAKDT 762
Db 719 TQAFMLTPPQELGPHSCHSFOTTGTLDEPWEVLATQFCLRESEDSSETOPFDTHLEA----774
Qy 763 QONGP-KPATLIGADGPPGP-----PTPRTSRAPIE-----PTPASEATGATPPPA 811
Db 775 --YGPCLSPRAIPGDQHPESPVHTPEMGIOGRQOTVDKMGWIKPETAERVG-----PER 828
Qy 812 PRSPSPAPPPVVP-----KEEKEEETAAAPPE-----EGEBQKPPA-AE 849
Db 829 GPLERETKLLPERQTDVTGEBELTKGQDREQKOLLARDTQOESDKNGESASPERDRE 888
Qy 850 ELAVDTGKAE-----PVKSECTEAEAGKAGKDAEAB-----ATAEGALKA 893
Db 889 SLKVEIETSEIOEKQVQKTLPSKAPEREVERPVANRECDPAELEEKVPKVLERDITOR 948
Qy 894 EKEGGSGRATTAKSGAPQD-----SDSSATCSADEVDEAGDKNRLLSPR-----PSLL 945
Db 949 GEPEGGSQDQKQASSPTPEPGVAGDLPGITSAPVPVSGSGGGRGSPVSPRRHKGSGILL 1008

Qy 946 TPTGDPANASQKPLDLKQKORAAAIPIPIQVTKV---HEPPREDAAPTKPAPPAP- 1000
Db 1009 NCKMPAEKAS-----RIRAA-----EKVSRGDQESPDACLPTTVEAPAPQ 1051
Qy 1001 PPONLOPESDAPQOQSGSPRGKSRGPAPPADKEAFAAEAKQLPGDPPCWTSGL-PF-PVP 1058
Db 1052 KPLNSQSKHLAPPLPLLSPLLSIKPTVRKTRQDGSQEAPEAP-----LSSELEPHPKP 1106
Qy 1059 PREVIKASHADPSAFSVAPPCHPPLPLGLHDTARVLP-PTISNPP 1107
Db 1107 KIRTRKSSMTFPP-ATSAPEHPST-----STAQPVTKPTSQATRSNTNSSVKTPP 1161
Qy 1108 LISSAKHPSVLEROIGAIQSGMSVOLHVPYSEHAKAPGVPTMGLPLPMDPKKLAPFGV 1167
Db 1162 VVPTA-----FELQPSSTQDQVTS-----EPTSQV 1187
Qy 1168 KQQLSPRGOAGPPESLGVPTQAEASVLRGTALGSGVGSITKGIPTSTVPSDSAITYRG 1227
Db 1188 TRGRKS-RSSVKTPETV-VPTALEQ-----PST-----1214
Qy 1228 SITHGTPADVLYKGTITRIIGEDSPSRDLGRSDSLPKGHVYEGKKGHVLSEYEGMSVT 1287
Db 1215 -----STDRPVTSEPTSOATRGRKN-----RSSVKTEP-----PVV 1245
Qy 1288 QCSKEDGRSGSPPHETAAPKRTYDMMEGRVGRGAISSASIEGLMGRAIIPERHSPHLKE 1347
Db 1246 PTAPELQPSSTQDQVTSBP--TYOATRGRKNRS-SVKTPPEPVPTA--PE-----1291
Qy 1348 OHIRSGITGIPRSVVEAQEDVLRKAEKLLKEEGTPPPPPSRDLTEAVKTOALGPLKL 1407
Db 1292 ---LRPSTSTDRPVT-----PAKTSR---TTSRSTN-----1316
Qy 1408 KPAHEGLVATVKEAGRSIHEIPREELRHTPELPLAPRLKEGSIQTGTPLYDTGASTTG 1467
Db 1317 -----MSSVKTPETVPTAPELQISTSDQVTKPT-----SRTTR 1353
Qy 1468 SKHDSRLIGSGPRTPPVHPPLDVMADARALERACRYEESLKSPPGTASSGSGSIARGAP 1527
Db 1354 SRTN--MSSVKNPESTVP-----1369
Qy 1528 VIYPLGKPKQSPLYTIEDHGAPFAGHLPGRGSPVTMBREPTRLQEGSLSSSKASQDRKITS 1587
Db 1370 -IAPEL-----PSTSTSE-----OPVT-PEPT-----SRATEGRK-- 1397
Qy 1588 TPRIAKSPHSTVPEHHHPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTSIPRGIPLDAAA 1647
Db 1398 -NRSSGKTPTLVP-----TAPK---LEPST 1419
Qy 1648 AYILPRHLAPNTYPHLYPPYLIRGYPDTALENROTIIINDYITSQOMHNTATAMAQRA 1707
Db 1420 STDQF--VTPEPT-----SQATRGRTRNSSVKTPETVVP--TAPELQVSTSD-- 1463
Qy 1708 DMLRGLSPRESSLALNVAAGPRGIIIDLSQVPHLPVLVPTPGTPTATMDRLAYLPTAQOP 1767
Db 1464 ---OPVTPEPTSOA-----TRGTDSSVK-----TPETV-----VPTAPEL 1497
Qy 1768 FSSRHSSPSLPGOPHTLTKPTTTSSSERERDRDRDREREKSLITSTTVEHAPIW 1827
Db 1498 QASASTDQVPT-----SDPTSRTT-----RGRKNRSVKTPETVVPAAPPEL 1538
Qy 1828 RP--GTEQSSSGSGSGSGSSSRPASHAHQHSPISPRTODALQORPSVLHNTGMKG 1885
Db 1539 QPPTSTDQVPTPEPTSRATRGRTNK-SVKTPESIVPIAPELQ-----PSTSRNQ-----1587
Qy 1886 IITAVESEKPTVLR-----STSSSPVRPAATFPFATHCPLGLGLDGVVPTLMEVLLPKE 1941
Db 1588 LVTPEPTSRATRGRTRNSSVKTPPEVPTAPEPHT-----TSTDQVPTPKL 1634
Qy 1942 APRVARPERPADTCHAPLAP--PARSGLEPASPSKSGSEPRPLVPVPSGHATIARTPA 1999
Db 1635 TSRARRKTRNSSVK---TPKPEVFAASDLEPFTTQDSVTPPEIAQ--GGQSKTLRSST 1689


```
RESULT 80
Q14676
AC Q14676 PRELIMINARY; PRT; 2089 AA.
ID Q14676;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA0170.
GN KIAA0170.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Myeloid leukemia cells;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 3:17-24(1996).
DR EMBL; D79992; BA011487.1; -.
DR Genbank; HGNC:21163; MDC1.
DR GO; GO:0005622; C:intracellular; IEA.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR000253; FHA.
DR InterPro; IPR008984; SMAD_FHA.
DR Pfam; PF00498; FHA; 1.
DR SMART; SM00292; BRCT; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS00172; BRCT; 1.
DR PROSITE; PS00006; FHA_DOMAIN; 1.
KW Hypothetical protein.
KW SEQUENCE 2089 AA; 226690 MM; A628ECCT567EERS0 CRC64;

Query Match 3.2%; Score 419.5; DB 4; Length 2089;
Best Local Similarity 19.6%; Pred. No. 1.1e-11;
Matches 425; Conservative 256; Mismatches 762; Indels 725; Gaps 106;

QY 486 SYRRRGKQQQQQQQQQ---QQQQQQQQPMRPSQOEKDEKEKEAB-----KE 532
D 486 SYRRRGKQQQQQQQQQ---QQQQQQQQPMRPSQOEKDEKEKEAB-----KE 532
QY 235 SAARRGATVEAKQSEAVVTVIQLEKQDQPLVK---ERDNDTKVKRGAGNGVWVAGVILE 290
D 235 SAARRGATVEAKQSEAVVTVIQLEKQDQPLVK---ERDNDTKVKRGAGNGVWVAGVILE 290
QY 533 EEKPEVENDKEDLLKEKTDTS-----GENDKE-----AV 564
D 533 EEKPEVENDKEDLLKEKTDTS-----GENDKE-----AV 564
QY 291 RSQPPGEDSDTDV-----DDSRPPGRPAEVLHRAQPFPGFIDSDTDAEERIIPATPVVI 345
D 291 RSQPPGEDSDTDV-----DDSRPPGRPAEVLHRAQPFPGFIDSDTDAEERIIPATPVVI 345
QY 565 ASKGRKTANSQGR-----KGRITRMA-NEANSEEAITPQQ---SAAELASMEINSSR 614
D 565 ASKGRKTANSQGR-----KGRITRMA-NEANSEEAITPQQ---SAAELASMEINSSR 614
QY 346 PMKRRKIPHGVTGRGAPGLAHLOESQAGSDTDVEEKGKAPQAVPLEKSQASWINSD-- 403
D 346 PMKRRKIPHGVTGRGAPGLAHLOESQAGSDTDVEEKGKAPQAVPLEKSQASWINSD-- 403
QY 615 WTEBEMETAKGLELHGRN-----WS-----AIARMVGSKTVSOCKNFYFNKQRQ 660
D 615 WTEBEMETAKGLELHGRN-----WS-----AIARMVGSKTVSOCKNFYFNKQRQ 660
QY 404 -TDDEEVSAALTLAHLKESQPAIWNRAEDMDPORVLLORSQTTE-----R 451
D 404 -TDDEEVSAALTLAHLKESQPAIWNRAEDMDPORVLLORSQTTE-----R 451
QY 661 NLDEILOQHLKMEKERNARK---KKKAPAAASEEAFPPVVE--DEEMEEASVSGN--- 713
D 661 NLDEILOQHLKMEKERNARK---KKKAPAAASEEAFPPVVE--DEEMEEASVSGN--- 713
QY 452 DSDTDVEEELPVE--NREAVLKDHTKIRALVRAHSEKQDQPPFGSDSDSVEADKSPGIHL 510
D 452 DSDTDVEEELPVE--NREAVLKDHTKIRALVRAHSEKQDQPPFGSDSDSVEADKSPGIHL 510
QY 714 -----EEMVBEAEALHASGNEVPRGEGSPATVNNSSDTSIPSP----- 754
D 714 -----EEMVBEAEALHASGNEVPRGEGSPATVNNSSDTSIPSP----- 754
QY 511 ERSQASTTVDINTQVEKEVPPGSAIMHIKKHQV-----SVEGTNQTDVKAQVGPAGKLL 563
D 511 ERSQASTTVDINTQVEKEVPPGSAIMHIKKHQV-----SVEGTNQTDVKAQVGPAGKLL 563
QY 755 -----HTEAAKDTGQNGKPPATLADGADGPPGPPPTPRRTSRAPISPTASEATG 804
D 755 -----HTEAAKDTGQNGKPPATLADGADGPPGPPPTPRRTSRAPISPTASEATG 804
QY 564 VVSLEEAWPLHGDCEQD-ABEGTSLTASVAD-----VRKSQLPAGSDAGAEWAA 612
D 564 VVSLEEAWPLHGDCEQD-ABEGTSLTASVAD-----VRKSQLPAGSDAGAEWAA 612
QY 805 APTPPPPAPPSAPPPVVPKEKEEETAA--APPVEEGEOKPPAAE---FLAVDTGKAE 859
D 805 APTPPPPAPPSAPPPVVPKEKEEETAA--APPVEEGEOKPPAAE---FLAVDTGKAE 859
QY 613 A-----VLKQERAHEVGAQGGPPVAQVEQDLPISRENLTDLVVDT----- 652
D 613 A-----VLKQERAHEVGAQGGPPVAQVEQDLPISRENLTDLVVDT----- 652
```



```
Db 1417 STSDQPVTEPTSQATGR-----TNRSSVKTPTVWTAPELOPSTSDQVTPTEPT 1470
Qy 1913 F-----PPATHCLGTLGVPTLMPEVLLKPEARVPRPRPRADTGHAF 1959
Db 1471 SQATRGTRDSSVKTPTVWTAPELOQASTDQVTSPTSRTRGRKRRSSVKTPTETV 1530
Qy 1960 LAKPPARSGLEPASPSKGSBPRPLVPPVSGHATIAETPAKNLAPHASDPDPAPPASAS 2019
Db 1531 V---PAPELOPPTSTD-----RVTPEPTSRATRGTRNSSV-----KTPESIVPIAPEL 1578
Qy 2020 DPHEKTKQSKFFSQEELRSLGVHGSYSYSGVEPVSPVSSPBLTHDKGLPHLELDK 2079
Db 1579 QPSTSRNLQVTEPTSRATRCRTNRSSVKTPEPVVPTAPEPHPTTSDQVTPKLTSRAT 1638
Qy 2080 SHLEGEALPKPGVVKLGGEAAHLPHLRPLPES-----QPSSSPLLQTA-----PG 2125
Db 1639 RRKTNRSSVKTPEKFEV---PAASDLEPFTTDDQSVTPBAIAQGGOSKTLRSVRAMPVPT 1696
Qy 2126 VKGQRVVTLAQHIS-EVITO-----DYTRHHPPQQLSAPLPAPLYSFP----- 2167
Db 1697 TPEQSPVTTQDIPSPETIPQSCIKQRAAGNPGSLAAPIDHKPCSAPLEPKSQASRNQ 1756
Qy 2168 --GASCPVLDLRPPSDLYLPPPHGAPARGSPHSGKSGSPENPKTSVLGGGSDGIEPV 2225
Db 1757 RWGA-----VRAAESLTAIDEP-----ASPQLLETPIHASQIQKVEPA 1794
Qy 2226 SPPEGMTEPHSRANVPLLYRDGEOTEPEPBMGS-KSPGNTSQPPAPFSLKTESNAWK 2284
Db 1795 -----GRSRFTPELQPKASQSKRSLATMDSPPHQKQP----- 1827
Qy 2285 SKQEIKNKLTNRNPEYNISQGTIEFNMPAITGTGLMYRSQAVQEHASTNMGLEA 2344
Db 1828 -QREVSQK-TVIIKEEEDTAEPKKE-----EDV 1856
Qy 2345 IIRKALMGKYDOWEESP---PLSANAFNPLNASLPAAMPITAADGRSDHLLTSPGG-- 2399
Db 1857 VTPKPKRKRQDAEEENPNRPSRLRTKLQNESTAPKVLFTGVVDARGERAVLALGSL 1916
Qy 2400 -GGKAKVS 2406
Db 1917 AGSAAEAS 1924

RESULT 81
Q17343 PRELIMINARY; PRT; 6994 AA.
AC Q17343
DC Q17343
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE UNC-44 ankyrins.
GN UNC-44.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=95263663; PubMed=7744957;
RA Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,
RA Boontrakulpoontawe P., Jeyaprasath A., Hedgecock E., Wheaton V.I.,
RA Sobery A.;
RT "An ankyrin-related gene (unc-44) is necessary for proper axonal
RT guidance in Caenorhabditis elegans.";
RL J. Cell Biol. 129:1081-1092(1995).
RN [2]
RP REVISIONS, AND SEQUENCE OF 6126-6994 FROM N.A.
RC STRAIN=N2;
RA Otsuka A.J.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39847; AAB41827.1; -.
```

```
DR EMBL; U21733; AAB38384.1; -.
DR HSP; P42773; 11HB.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000488; Death.
DR InterPro; IPR001360; Glyco_hydro_1.
DR InterPro; IPR000906; ZU5.
DR Pfam; PF00023; ank; 24.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00791; ZU5; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 21.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS00088; ANK_REPEAT; 22.
DR PROSITE; PS02297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00017; DEATH_DOMAIN; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 2.
KW ANK repeat; Repeat.
SQ SEQUENCE 6994 AA; 775364 MW; 90CB449925D9923D CRC64;

Query Match 3.2%; Score 419; DB 5; Length 6994;
Best Local Similarity 17.5%; Pred. No. 5.5e-11;
Matches 516; Conservative 371; Mismatches 1020; Indels 1040; Gaps 129;

Qy 16 ERYVPP-----HSLSYV-----QIARTHTDVLGLLEVQHHSRDYASHL--SPG 56
Db 3169 DPFSFAPSHPVSHVSGIHASESPVAQQEIPQTRFEHDSFAAQY-FHEEYEVSHVTEQA 3227
Qy 57 STIQPQRRP----- 70
Db 3228 PLUTEQQHQFSGDSGEGLSKVLGFAKAGVAGVVAAPVALAAVGAAYDALKK 3287
Qy 71 EFQGNERSQELHLRPESHLYLPELKGSEMEFTESKPRLELLPDLLR----- 119
Db 3288 DDDEEDQERESLQRQERSIDSPH--ASEQSIIEEHERFEESFPVSEKHHVTETTTTT 3345
Qy 120 -----PSPLLA-----TGOPAGSE-----DLTKDRSLTKLEPVSPSPPTDP 158
Db 3346 VTREYFDEHPLVSQETEGEGKSPASEKSLPHVVETTTTTVTREFDKNDSFVPSSEK 3405
Qy 159 EELVPPRLSKBELIQNMDRVREITWVEQOI SKLKKKQOOLEEAAKAPPEKPVSPPPP 218
Db 3406 EQE---PTTVSRVYETAEGDEPEHHYTTTITTTVTKEVIDDQOEMGDDDEP-KQESPVQ 3461
Qy 219 IESKRSVLQIIVDEN---RKAEEAARI----- 249
Db 3462 VETTTTTTSREYDNDDETREAGDSHITETTKTTTVTRFHFGEQPEETEETDEVELP 3521
Qy 250 POVE---LPLYNQPSD-----TROYNENIKINQAMRKKLILY 283
Db 3522 PKTEEDNVSEYSESSTSVSREVRPDPEPHILEITTTTTVTREYNEPEETVDDQDAPI 3591
Qy 284 FGRNRHARKQWKQFCORYDQLMALAEKKVRIENRRRAKSKVREYKEKQFPEIR-- 341
Db 3582 SFSQEH-QDDDSQASHDQHDR--ESPVESEKSVKHTTTTTTTTTVTQLYDDEASEIRGE 3638
Qy 342 -----KORELQERMQRVSGRGS 359
Db 3639 SPVATESEHVSTKSDSESHVPSVIETTTTTVTREFYDDQDELQREDHTOSEERRS 3698
Qy 360 GLMSAARSSEHVESEII-----DGLSEONLEK-QMRQLAVIP-----P 397
Db 3699 --SIPTEETHEDSHLKEITTTTTVTREFYDPEENVKQLQDSQFSLSPSHVSEIYVP 3756
Qy 398 MLYDADQORI-----KFINNGLMADPMKVKYKDRQVNMNMSBOEKE 438
Db 3757 ESPVAKQOEIPQTRFEHDSFAAQYFHEDEYEHQVPTQAPLITTEQHQH-PESGEESDGE 3815
Qy 439 TREFKFMQHPKNFGLIASFLERKTVAECSVL-----YYVLTKN-----ENYKSLVR--RS 486
```


Db 4287 GQGESDQIPSEISITSENMDRETSSSPVQNRDEEFVLPALPYKQTEYGRVDSDHAP 4346
Qy 811 APPSPAPPPVVPKEEKEEETAAAPVBEGBE-----QKPPAAEELAVDTGKAEPPVKS 864
Db 4347 ASPAPASPSPI--QAYKQESQEAHLSQFOQRSSVSHEDSPAQYFHDNEDHSDSPVPS 4404
Qy 865 E---CTEAEAGPAGKDAEAAEATAEALKAKEKG--GSGRATTAKSSGAPQDSQSAT 920
Db 4405 DRAPLLTQQHQPSGESD--GEGFGSKVLGFAKAGMVGAGVVAAPVALAAVGAAYD 4463
Qy 921 CSAEDVDAEGGDKNRLSPSLTLTPGDPRANASPOKPLD--LKQLKRAAAIPIQ-- 977
Db 4464 ALKKDDDEDNDPEOKLKP-----SPERQVLNVPEDSEISELEYTSPSPTEK 4517
Qy 978 -----VTKVHPFPRED-----APTAPPAP----- 999
Db 4518 SESQCYTETVTTVTREYLDPPQSVTRSPSEHDISEQYAPESPVQDPYVVEKITTV 4577
Qy 1000 -----PPQNLQ-----PESAPQPGSGSPRG-----KSR 1024
Db 4578 IRQVHDEPPQIEEQTIPPEVTVLREVYESPEGDEPEQHYIETKTITITKEVHVVPVEED 4637
Qy 1025 SPAPADKEAFAAQAQLPGPPCWTSGLPVPVPREVKA-----SPHAPDPSAFSY 1077
Db 4638 VOISPVHSETSVSEKQ--LPAD-----EQLDFEVTESTTTATVTRERYEPEEVRPPSGSE 4691
Qy 1078 APPGHPLGLGHDARTVLPRPPTIS-----NPPPLISSAKHPSVLEROI 1122
Db 4692 ADDSHAPKYMETTTTTVTREYEVSEDEDEHQSOVORDESPAPSESVKH--VIEK-- 4747
Qy 1123 GAISQMSVQLHVPVSEHAKAP-----VGPVTMGLPLMDP-----K 1159
Db 4748 -TTTTVTREYEPEDSHSPVEDDVGFKVTTTTTTTTVTHEPEPEPPSDEHVSE 4806
Qy 1160 KLAPSGYKQQLSPRG-----QAGPPESLGVPTAQEASVL 1195
Db 4807 RYASGSPVSEDSREIETTTTTVTRHFELEDQDQHVHVESYASGSPFVPSKSEVE 4866
Qy 1196 RGTALGSPVPGSITK-----GIP-----STRVPSDAITYRGSITHTGPADV 1237
Db 4867 R--VIETTTTTVTRHFEHEDDIPVETSHDDPAASVSEEDV-----HGQIQT 4917
Qy 1238 LYKGITR--IIGEDSPRLDRGSDSLPKGHVIEYEGKKHVLSEYEGMSV-----TQC 1289
Db 4918 TTTTIVTRHVVPDE--EIDSGRMDELEK-----YSESPPVSEEDSSRVETTTTTTV 4969
Qy 1290 SKEDGRSSGPPHETAAKRYDYMMEGRVGRAISSASTIEGLMGRAPPERHSPHILKEQH 1349
Db 4970 SREHFEPEP--DHSHVVSQBY-----SAS-----GSPVPSK-----SVR 5004
Qy 1350 HIRGSITQGIPIRSYVVEAQEDYLREAKLLKREGTPPPPPSRDLTEAVKTOALGPLKLKP 1409
Db 5005 VIETTTTTVTRHFEDEEDYIPSESRSHDDGITDQHVPSQS----- 5047
Qy 1410 AHEGLVATVKEAGRSIHPIREELRH--TPELFLAPRLPKESITQGTPLKYDTGASTTGS 1468
Db 5048 -----FVPSSEDDHVTHEQIIRDDPIDE--QIVESHKEYSESSIPSSQD 5089
Qy 1469 KKHVRSUIGSP--GRTPFPVHPLDMADA-----RALERACYEESL-----K 1509
Db 5090 STHVLETVMSTPITSERYDPEKDVESADDEIDSTAQYKSESPVQTEKSLILAKQ 5149
Qy 1510 SRPGTAS-----SSGSGIARG--APVIVPELG-----KPRQSPLTVED 1545
Db 5150 QESGDESDEGFGSKVLGFAKAGMVGAGVVAAPVALAAVGAAYDALKKDDDEEDQED 5209
Qy 1546 HGAPAGHPLRGSPVTRPEPT---PRLQE--GSLSSSKASQDKLTST-----PREI 1592
Db 5210 KEPLIGGFHKQDDPISQDEPSAVDSETPGATFVEPEEDKVITDSADSSVQDEPKIV 5269
Qy 1593 AKSPHSTVPEHHPIPISPYHLLRGVGVLDYRSHIPLAFDPTSPRGIPLDAAAAAYLP 1652
Db 5270 --FPVDSTPEHHSNDRESFESIVKSEG----- 5294

Qy 1653 RHLAPNPTYPHYPPYILRG--YPTDAALENROTIINDYITTSQOMHNTATAMA--ORADM 1709
Db 5295 -----PYIVESTDYAQTSAEPR-----ISSPVHSDAGDASSFKRPDES 5332
Qy 1710 LRGLSPRESSIALNYAAGPRGIIILSQVPHLPVLVPPTPGTPATAMDRLAVLPTAPQDPS 1769
Db 5333 VTGDEKNA-----IPETSETDAPV-----IDSEYEFPN 5362
Qy 1770 SRHSSPLSPGPGPHTLTKPTTSSSERDRDRDRE-----RDRDREREKSLITSTT 1819
Db 5363 NRDRQRISSP-----AHSDEDEDNAEVIDSEFYRHSEQONNEEDPSIV----- 5406
Qy 1820 TVEHAPIWRPTEQSSSGSSSGSGGGSSSRPASHS-----HAHQHSPISPR---TQDA 1870
Db 5407 -----ESGEYISSGHG---SPRPPEDSITTTTTLNVHHEPAALPEPEVDEEE 5449
Qy 1871 LQQRPSVLHNTGK-----GIITAVEPSKPTVLRSTSSPVPRPAATFPFPAHCPG 1923
Db 5450 LEQERSIIESEYKTSPLPPTSVTVEHVPEABIKYRTTSTP--IVTTVSSEH----- 5502
Qy 1924 TLD--GVYPTLMPEVLLPKEAPRVARPERPRADTG-----HAPLAKP-PAR 1966
Db 5503 -LDRNQEPYVVESEYTRASP--LGPERPESPSPLPREDDSHVIESHYTSSPVSE 5559
Qy 1967 SGL-----EPASSPSKGSBPRLVPPVPSGHATIARTPAKNLAPH--HASPD-- 2010
Db 5560 DSVKHVIEKTTTTVTVEERYEPEDSHSPVPSDDV--HGFVKTTTTTTTTVTHEHFEPEDH 5617
Qy 2011 -----PPAPASADPHREKTQSPFISQIQLSLGSLVH---GSSYSPEG 2052
Db 5618 TSDEHVVESERYASGSPVPSSENSRVTTTTTTTTVTRHFEPEDDQDQHVHVESQBYASG 5677
Qy 2053 VEPVSPVSPSLTHDKGLPKHLE-----ELDKSHLEGLRKPQKQPVKLGCEAAHLPLR 2107
Db 5678 -----SPVPS-----EKSVKEVIETTTTTVTRHFEHE--DEIPTIV-----ESSH 5717
Qy 2108 PLPESQSSSPLLQATPCVKGHQRVVTLAQHISEVITQDYTRHH---PQQLSAPLAPPLY 2164
Db 5718 --DDQAASS--VPSEEDVHQIQTT-----TTTTVTRHVVPDEIDSGRMDELE 5764
Qy 2165 SFGASCPVLDLRPPS-----DLYLPPDHGAPARGSPHSEGGKRS 2207
Db 5765 KYSSSES-PV-----PPEDSSRVETTTTTTTFIREHFEPEDDHSHVVGVSQBYASG--SP 5816
Qy 2208 EPNKTSV 2214
Db 5817 VPSEKSV 5823
RESULT 83
O60382
ID O60382 PRELIMINARY; PRT: 1791 AA.
AC O60382;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein KIAA0324 (Fragment).
GN KIAA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Temer J., Meincke L., Longmire J., White S.,
RA Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT "Sequencing of Human Chromosome 16p13.3";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

Db	862	AGGSVTTTTTTITTKSSPSAPASP---TTAAPTSPASSAPPDNKLSQVYTTTTKPGD	918
Qy	807	---TPPPAPSPSPAPPPVVKKEEETAAAPPVEEGEEKPP---AAEELAVDTGKAEEP	861
Db	919	IFSLPPTPTPTINNEPTLTTRNTTTTTTTTTTTATSDNLQNHKPSAIEPATDTD--SOP	976
Qy	862	VK---SECTEAEAGPA-----KGDAAEAATA-----EG	889
Db	977	IRKVKLSANEAKVVEEACVRYQYQLNGENETPETPSSCTPANKPKHQMRPHDPEP	1036
Qy	890	ALKAKEKGG---SGRATTAKSGAPQDSDSSATCSADE---VDEAEGGDKNRLLSPR	941
Db	1037	QLRSSKSPSVEPQVQREITTFEGRRVSQDRE---ISIDELIJIETSGAGP---SPK	1088
Qy	942	PSLLTPTGDRANASPOKPLDLKOLKORAAAIPPIQVTKVHBPREDAAPTKAPAPP	1000
Db	1089	IPS-----PRAQ--SPKQPATRSQ-----SPEKQPRATP	1116
Qy	1001	PPONLOPESDAPQPGSSPRGKSRSPAPPADKEAFNAEAKLPGDPPCWTSGLPFPVPPR	1060
Db	1117	---ROSPEKEQAFKQTHEVTPSQEKFPRARSPEKTPG---WTO--FOVSPRQ	1164
Qy	1061	EVIKASHPAPPSAFSAPPGLPLGLHDRTARVLPRLPPTISNPP--PLISSAKHPSVL	1118
Db	1165	SPEKQLPRAQSEK---VPAVRQSVSPRSQEKQIPDKPTDQOGPLPRISPRQSP---	1218
Qy	1119	ERQI-----GAISSQMSVOLHVPVSEHAKA	1143
Db	1219	EKQLPKDVPQKSRQSPKDLTNQORREBEIIFRSTITTTTQKRTTNLNEEFITNERDNQ	1278
Qy	1144	PVGPTMGLPLPMDPKKL-----APFSVKQEQLSPRGOAGPESLGVPTAQASVLRG	1197
Db	1279	PISEKKQIPANAEPTKSETIESPDGGFPKSKTTEVEAQ---PEVKESPIYRKGLJRR	1395
Qy	1198	TA-----LGSVPGSGITKIPSTRVPSDSAITYRGSITHGTPADVLYK-----	1240
Db	1336	ETFEDECRQILGMEEDGD--TQTYTYERNNEQEDV---NVSHTTIETIQVIEDCPNDDE	1391
Qy	1241	-----GTITRII-----GED-----SPSLDRGRED--LPKG---	1266
Db	1392	DDKRRVTYTVVTRTPQIKVBEELFVDVTEADVEILVNFPSKSPKEEDSPKYPKGPE	1451
Qy	1267	-----HVIYEGKKG-----HVLSEGGMSVTQCSKEGDR-----	1295
Db	1452	PKSPNDQRIPIPKKQSPVQFKTEETPRYPQEQERYPKPETSQYKESPRNPKEADAE	1511
Qy	1296	-----SSSGPPHETAAPKRTYDMMEGRVGRA-----ISSASIEGLMGR	1333
Db	1512	TININEETIVITKEGSKSPSRWSPSPER-----RVPKSQQPPSPASPSVPSVGR	1564
Qy	1334	ATPPERHSHLHKEQHHRGSIQTQIPRSYVEAQEDYLREAKLLKRGTPPPPPPSDL	1393
Db	1565	KIPNEVESNFVTEKIIDCRG-----KTWVE-----KISQRPRTSPPTPKNT	1607
Qy	1394	TEAYKTQALGP--LKLPAHEGLVATVKEAGRSIHBIPRELHRTPELPLAPRPLKE---	1448
Db	1608	KFSQKIPERVPETESEPEKDSSETKKTSVSVTETERNSTRTKTKQLPLKEPOSK	1667
Qy	1449	-----GSITQGTPLKYDVTGASTTGSKKHVDVRSILIGSPG-----	1481
Db	1668	VPAGKSPRKDSLTKRKRDSLVBETRIITTTTTTTRQGRKPSDTN---GSPSFKDLRLSSPR	1724
Qy	1482	-----RTFPVPHL-----DVMDADALERACVYESLKRPGTASSGSGSIARGAPV	1528
Db	1725	KQKTSPOQRTPTPAQTRNPDVDVGDSSS-----PDASPTRVGNERRRRNSNIVHTEI	1778
Qy	1529	IYPELGKPRQSLPTLTDHGCAPFAGHLP---RGSPVTMBE---PTPRLQEGSLSSKASQD	1582
Db	1779	IDHMAP--KSPKT--ERRSQGTGNVPSPIRKLPVTERKESAPVRV-----TERRDKA	1828
Qy	1583	RKLT-STPREIAK-----SPHSTVPHEHHPHIPSYPEHLLRGVSGVDLYRSHIPLADPTSI	1637
Db	1829	EKVTSTSENIIKMSGTHKPHPEMSLKPGGDRSRPSKCTTKITINLSEORINTATD----	1884
Qy	1638	PRGIPLDAAAAYVLPRLAPNPTYPHLYPPYLIRG---YPTDAAA-----	1678
Db	1885	IEGVIIIDIOQA--KSSREPSDRIVTPVPAELETGKPRIPDVVOEPDDDEPRKPKQVTNIP	1943
Qy	1679	--LENRQTTIINDYITSQOMH-----HNT	1699
Db	1944	IFEEESQTYVGCQIS--ELHSSNGIEVDILDNPTVEAPKSLDYVNTPTDDESLLSVHEK	2001
Qy	1700	ATAMAGRADMLRGLSPRESS-----LALN-----YAAGRGILDLIS	1735
Db	2002	VSRFTSHSAEKVK--EPKVSAPFSREDFVNAKIPENDDCLLSNQKVDKFLRTAENVIRPT	2059
Qy	1736	QVPHLPVLPVPTPGTP-----ATAMDLAYLPTAPOFPSSRHS	1773
Db	2060	SLPSRPEI--ERPELBEIDELRDCTLSVSKVHKFIDTAEKLA--PTMPO-----K	2109
Qy	1774	SSPLSPGGPHTLTKPTTTSSSERERDRDRDR-----BREKSILSTTT	1820
Db	2110	SPRLVANIHERHISR--QSEPERELDESEPELDRDVEDDDQTSQLETEEBEITQVTIK	2166
Qy	1821	VEHAPITWRPOTEGSSGSSGSSSRPASHAHQHSPISPRTODALQORPSVLHN	1980
Db	2167	KETLKBFKQOTKETRETRRDSKAEPEKLOKSPQTKVKEESARVPKYQAKVOK-----	2220
Qy	1881	TGMKGIIITAVEPSK-----PTVLRSTSTSPVRPAATFPFATHCPGLGGTLDGVYPTLMEP	1935
Db	2221	-----VSQWEPKKQOREPKV--TQKETPLFPKKQ-----PLSKVKD-----EP	2257
Qy	1936	VLLPKAPRVAR-----PERPRADTGHAFILAKPPAR--SGLEFPASSPSKGSBERPL	1984
Db	2258	EKNKREPKVPQESQTKLKEEPERTKTKTPOKEPRKEPLRQSEDEPEFSPSEEFDDDEPL	2317
Qy	1985	VPPVSGHAT-----IARTPA--KNLAPHASDPDPAPPASASDPHREKTSQSPESI	2033
Db	2318	-PMTKHTHTTAEMKRQKDIILNRPSVFGQRTPEKSSSTTSP-----TKLNGTRGRSPS	2370
Qy	2034	QEL---ELRSLGVHGSYSPEGEVPVSP--VSSPSLT---HDKGLPKHLELDK-----SH	2081
Db	2371	TNLITEKRSYRQVNTVNSKPGTKRTKTPSANSQAQPPPKTTTSIKRMEQISQGSWVQD	2430
Qy	2082	LEGELRPQGPVKVLGGEAAHPLHPLPESQ---PSSSPLLOTAPGVKHQHVTLAQH	2138
Db	2431	VDVDVEVVGAPPS-----HISEKPOGKSPSTSSRSLSRSPSPSRSTSTNLNT	2481
Qy	2139	ISEVITODYTRHHPOQLSAPLAPALYSPFCASCPVLDRPPSDLYLPPDHGAPARGSP	2198
Db	2482	TSTNTTT--TTEHPSTTKTTTPKPTSTNPKDBPEI-----IP	2517
Qy	2199	HSEGGKRSPEPNKTSVLGG-----GEDGIBPVSPPEGMPTEP-----GHRSASVYPLL	2245
Db	2518	IESLTEKSIITTYTNTTGRNVASRRNVPEPVHETHVDSEPTGRRPSPYMDHTKSSL-EHI	2576
Qy	2246	YRDEQTEPGRMSKSGNTSQPPAPFSPKLTESANSAMVSKKQKQKINKLTHNNEP---	2302
Db	2577	RROSLINKSHYRSKSMEDDSP---VEPRNPNS---SVKFDVPRKSSRSGADEPRKT	2628
Qy	2303	-----EYNISQPGTIEFNMPAITGGLMTYRSQAOE-----HASTNMG	2341
Db	2629	SLKGDSDSDELEIEIFDLQRLKLE--LETVASIEMRRIRRAQMRILIRKMNINAGTTTT	2687
Qy	2342	LEAIIRKALMGKYDQWEEESPPLSANAFNPLNASASLPAAMPITAADGRSDH-----TL	2394
Db	2688	ITTTTSTTPCK-----SSPLPKIRRDQSPAGAAEVKTEVTTTTTSSRQOQOQVEQVDS	2741
Qy	2395	TSPGGGKAKVSGRPSR---KAKSPA-----PGLAS--GDRPPSVSSVHSEGDNR--	2441
Db	2742	TTTAPAGKTSHPGKPPVKPRERSASPAQKRIRISPPGQSPDRSTTTTTTKVTITSTTRGA	2801
Qy	2442	-RTPLTRVWEDR-----PSSAGSTP	2461
Db	2802	PSKPAQGPINADRSKVLKGHATVQTNGSTP	2832

Db 106 RLASTNSVPGVPLPSSTKEKGVAKTNKASQVNS--TTHPASAKAVAHILSGRSPRKAG 164
Qy 819 PP--VVPKEEBE-----TAAAPV-----BEGEOKPAAAEELAVDTGKAEEPVKS 864
Db 165 PSANTILVSETEEGSVPALGTTAKPGMASANQADSSSETSSSDETDVEKASEKIVQA 224
Qy 865 ECTEAEAGPKAGKDAEAAEAGALKAEKKEGGSGRATTAKSGAPQDSATCSAD 924
Db 225 ---KAASGPVKGTGKATPAPP-----KAGPS--AAQAKTEKPKEDSDSS---EED 269
Qy 925 EVDAAEGDKNRLLSPRSLTTP---TCDPRANASPKQLDLKQLKQBAAIPIQVTKV 981
Db 270 SDSSEEPAAKTLQVKPSGKTQVKAASAKESPRK-----GVPPVPPGKV 317
Qy 982 -----HEPPREDAAPTAP-----PAPPPQWLQ-----PESD 1010
Db 318 GPAQAKKAGAGEEDPDSTSESEEAFTAVPTTRSPVQAKFSGQNSQVRTASGPVK 377
Qy 1011 APQO--PGSGPRGSRSPAPPADKEAFAAEAKLPGDPDPCWTSGLPFPVPPREVIKASPH 1068
Db 378 PPQKAGPAATPVGQEE-----DSSESSSEESDSEGAAPAQA-----KSSGK 419
Qy 1069 APDPSAFYAPPGHPLPLGLHDTAPVLPPTTNSPPPLISSAKHPSVLEROIGALSOG 1128
Db 420 IPQVRAASGPARG-----PPQKAGPAATQAKAK-----MSKDDSESEE 458
Qy 1129 MSVQLHVPVSHAKAPGVPTMGLPLPMDPKLAPFSGVKQEOQLSPRQAGPPESLGVPT 1188
Db 459 ESES-----ESEAAMAPVQAKSTVTPQYKASPKGT---PITPAPAKPPVVRGTGA 510
Qy 1189 AQEASVLRTALGSPGGSITKGPSTPRVPSDSAITYRGSITHTGTPADVLYKGTITRIIG 1248
Db 511 PWKARAEAS PACASSP--AMARGAQKPEASSS-----541
Qy 1249 EDSRRLDRGRDLSLP---KGHVIYEGKKGHVLSYEGGMSVTQCSKEDGRSS-----GP 1300
Db 542 EESES-----BEETAPAPAAGQAKPVGK-----GVTVKAASPTPTKPSGOGTALAP 587
Qy 1301 PHE---TAAPKTYDMMGRVGRATISSASIBGLMGRAITPPE---RHSPHLKQHHIRG 1353
Db 588 POKAGPATAPVKTVEQEDSSSESESESESESESESESESESESESESESESESESESE 647
Qy 1354 SITOG---IPRSYVBAQDYLRREAKLLKRGTPPPPPPSR-DLTEAYKTOALGPKLKP 1409
Db 648 TSAKGAASAPGVGSA-----AVQIKQESPAKVKEPPARTSQSNVSVRGQAGV---P 696
Qy 1410 AHGULVATVKAGRSIHIEIPRELHPTPLAPPLKEGSIQTQTPLYKDTGASTGSK 1469
Db 697 AVGKAVATAAQA-----QPGPVKNPEEDSESESE 725
Qy 1470 KHDVRESLIGSPGRTPPPVHPLDVMADARALERACVYESLSKSRPGTASSSGGSIARGAPVI 1529
Db 726 ESD-----SDGEAPTVPK-----SGKTQVGTASAPSGLGRKGATA 763
Qy 1530 VPGLKPKRQSPLYTDHGAFFAGHLPRGSPVMTREPTPLRQGLSSSSKASQDRKLITSTP 1589
Db 764 AP---PKTGAVATAQAKP-----EEDSESSSEESDSE-----794
Qy 1590 REIAKSPHSTVPEHHPHIPSP-----YEHLLRGVSGVDLYRSHIPLAFDPTSIPIGP 1642
Db 795 -----EEAPQVKPGIKPPQTKASLLKGVSGT-----P 822
Qy 1643 LDAAAYVLPRLAPNPTVPHLYPPYLIRGVPDTAALENRQTIINDYITTSQMHNTATA 1702
Db 823 ASAEASTTVDSNAPQKARP--APP-----AKEGSGSKTAKS 857
Qy 1703 MAQRADMRLGLSPRESSIALNYAAGRGIIDLSQVPHLPVLVPP-----TPGTPTATA 1754
Db 858 KAQAP-----APPEKN-----AEGSSSD--DELPATQVIKPPPLIFVDPNKRSPAGPAAT 905
Qy 1755 MDRLAYLPTAPOPFSRSHSSPLSPGCTHLTKPTTSSSRERDRDRDRDREREKSI 1814
Db 906 ATQ---APAADTPRKAQAS-----TARSSSESEDE-----V 937

Qy 1815 L-----TSTTTVEHAPIWRPCTEQSS--GSSGSGGGGSSRRPASHAHQHSP 1862
Db 938 IPATQCSVPAGRINVITPLTAHP--RPAVRASTVGASGEASGRVSEK-----KQEAP 988
Qy 1863 ISPTQDALQOORSVLHNTGKGIITAVEFSKPTVLRSTSTSPVRPAATFPATHCPLG 1922
Db 989 ITQVTKENPAHLP--LTQAAKLVLAQKASEAQPPAARTPSSSGVDHALGTLF-----1038
Qy 1923 GTLDGVYPTLMEPVLLPKEAP--RVARPERPRAD-----TGHAFKAPKPARSGLBPAS 1973
Db 1039 -----VMSQITPVQAKMTNKLKAEVPAVERATAITPVGHPKAKASETSDSDSS 1089
Qy 1974 SPKSGSEB-----RPLVPPVSGHATIA-----TPAKNLAHHASDPDP 2012
Db 1090 GSSSGSEEDADGPKAHRVGPAPSATETILVEETABESSEDEVVAPQSLLSGVVTGPT 1149
Qy 2013 APASADPHREKTQSPFSIQELESRLSGHYGSSYPGEVPSVPSFSLTHDKGLPK 2072
Db 1150 LANSQAS-----KATRP-----DLNP--SASSTSIKDAPD 1179
Qy 2073 HLEELDKSHLEGELRPQPGPKVLGGEAHLPHLRPLPESQSSSPLLQTAPGVKHQHV 2132
Db 1180 GKQEVPEQHVAGTVSPKTS-----RREADATPQ--KRPKPKKEAGSP-----QA 1221
Qy 2133 VTLAQHISEVITQDTRHHPQOLSAPLAPLYSPFGASCPLV-----DLRRPPSDLYLPP 2187
Db 1222 STLA-----LQSDISR---RLISEPWPLNEAQVQASVWKVLTLLBOEKKAADAAKES 1272
Qy 2188 PDHGAAPARGPHS--EGGKSPSPKNTS--VLGGEDGIEFVSPPEGTWP--GHSRSVY 2242
Db 1273 NRKGRVGRKRLSGDQTAARVPKSKKKQLVAGGGGGAAG--SPEKALRTPKGAK-----1327
Qy 2243 PLLYRDGEQTE--PSRMSKSPGN-----TSOPPAFFSKLITSNAMVKSQKQEIKNKL 2294
Db 1328 ----RDGASDIEKKESKESFGLSEAKEKEGEPGTILKVGDOGNFKIKEKKSKDKK 1383
Qy 2295 NTHRNEPEYNISQGT 2312
Db 1384 KDEKKEKKKKKAKKASTK 1401
RESULT 88
Q966V1 PRELIMINARY; PRT; 2376 AA.
AC Q966V1
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE REXIN LI.
GN CAP OR RXN OR CG3451 OR CG18408 OR CG18409.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamazaki H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Willert K., Fish M., Nusse R.;
RT "Drosophila REXIN, a Novel SH3 Adaptor Protein of Axin and Arrow that
is Essential for Living in Late Stage Embryo.",
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: CONTAINS 3 SH3 DOMAINS.
DR EMBL; AB053478; BAB62017.1; -
DR Flybase; FBGN0033504; CAP.
DR InterPro; IPR00108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 3.
DR PRINTS; PR00499; P67PHOX.


```
QY 2259 SKSPONT---SOPPAFFSKLTESNAMYKSKQKQKINLKNTHNNEPYN-----ISQ 2308
:|||||:
Db 1097 AKSPIATPPVSTPIAKPPPIATPPVATTPIAKPPI---ATPPTANPPVANPPIAKSPIAK 1153

QY 2309 P--GTEIPNPAIT---GTGLMTYRSOAVQEHASTNGLAIIRKALMGKYDQWESPP 2362
:|||||:
Db 1154 PPIATPPVAMPISIATPPPIGKPPVATPPWAKPPVASPPIATPPPIIKPPV-----ATPP 1205

QY 2363 LSANAFNPLNAS---ASLPAAMPITAADGRSDHTLTSPGGGKAKVSGRPSRKAAP 2419
:|||||:
Db 1206 IT---KPPVATPPVATPIAKPPVA-----TSP-----ITPPVAKPPVTTT 1244

QY 2420 GLASGD--RPPSVSVHSEGCNRRTPLTNRWEDRPSAGSTPPYNPLIMRLQAGVMA 2477
:|||||:
Db 1245 PVATPPVVKPPIVTPPIATPPI-AKSPIA-----PPPIGTPPIAKPP-----VA 1287

QY 2478 SPPPPGLPAGSPLAGP 2494
:|||||:
Db 1288 TPPTATPPVATSPIAKP 1304

RESULT 93
Q9H4A3 PRELIMINARY; PRT; 2382 AA.
ID Q9H4A3
AC Q9H4A3;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Hypothetical protein.
DE
GN WNK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Verissimo F., Jordan P.;
RT "A novel human protein kinase gene, WNK1: cDNA cloning, expression and
RT genomic structure.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AJ296250; CAC15059.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR008271; Ser thr_pkin_AS.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot kinase; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein; ATP-binding; Kinase;
KW Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 2382 AA; 250755 MW; ADDA6C86554B7FCA CRC64;

Query Match 3.1%; Score 405; DB 4; Length 2382;
Best Local Similarity 19.6%; Pred. No. 6.8e-11;
Matches 513; Conservative 291; Mismatches 943; Indels 872; Gaps 114;

QY 147 PVSPPSPHTDPELELVPPRLSKELIQNMDRVDREITWVEQISLKKKQKQLEBAK 206
:|||||:
Db 99 PLSLPQSPIPAAPQSAPEPHRETV-----TATATSOVAQPPAAAAGEQVA 149

QY 207 PPEP-----EKVPSPPIESKHSRLVQIYDENRKAEEAHRILEGLGQVVELPLY 257
:|||||:
Db 150 GPAPSTVPSSTSKORPVQPSL-----VGSKEEPPPA 181

QY 258 NQPSDTROYHENIKINQAMRKLLILYFKRNHARKQWKQKFCORYDQMLEALEKKVERIE 317
:|||||:
Db 182 RSGSG-----GGSAAKEPQEESSQQQDDI-BELETKAAGMS 215
```

```
QY 318 NNRRRAKESKV-REYEEKQPEIRKOR-----ELQERMOSRVQSGSLSMSAARSE 369
:|||||:
Db 216 NDRGFLKFDIEIGRGSKFTVYKGLDTEITVEVAMCELQDRK-----LTSEORGF 265

QY 370 HEVSEIIDGLSEQENLEKQMLQAVIPMLYDAQOQRIK-----FINNMGLM-ADPMKYK 424
:|||||:
Db 266 KBAEAMLKGL-QHFNIVR-----FVDSWESTVKGKKICVLVTELTMTSGTLKTYL 313

QY 425 DR-----QVNMNMSQEKETFEKFMQHPK-----NFGLIA 455
:|||||:
Db 314 KRFVKMKIKVLRSCRQILKGLQFLHTFTPIIHRDLKCDNIFITGPTGSKVIGLDGL-- 371

QY 456 SFLERKTVAECVL-----YYLYTKKQENY-----KSLVRSYRRRRKSKSQOQOQO 501
:|||||:
Db 372 ATLKRASFASVIGTPBFMAPEMVEEKYDESVDVYAFGCMCMLEMATSEYPYSECNAAQI 431

QY 502 QOQOQOQOQP-----MPSRSQ-----BEKDEKEKEKEKEKEKEPEVENDKEDLLK 547
:|||||:
Db 432 YRRVTSGVKPASFDKVAIPEVKEIIEGCIRQNKDERYSIKOLLNHAFQOETGVGRVLA 491

QY 548 EKTDDTSGEDNDEKEAVASKGRKKTANSQGRKGRITRSMANEANSE-EAITPQQAELAS 606
:|||||:
Db 492 E-----DDGEKTAIKLWLRIEDIKKLKGKVKDNEAIEFDFDLERDVPDVAQ--- 538

QY 607 MELNESSRWTEEMETAKGLLEHGRNWSAIRMVSGSKTVSQCKNFYFYNYKKRQNLDEIL 666
:|||||:
Db 539 -EMVESGVYCEGDHKTWAKIKDR-----VSLIK-----RKREORQLVR 576

QY 667 QOHKLKMEKERNARKKKA-----PAAASEEAPFPVVEDEMEASG 709
:|||||:
Db 577 BEQKKQKQESSLKQOQVEQSASQTGIKQLPSASTGIFTASTTSASVSTQVEPEPEA-- 634

QY 710 VSGNEEMVEEAEALHASGNEVPRGECGPGPATVN--NSSDTEISPSHTEAAKTGQNGP 767
:|||||:
Db 635 -DQHQQLOYPQPSISVLSDGTVDGSGQSSVFTESRVSSQQTIVSGSQH-EQAHSITGVPG 692

QY 768 KPPATLGADGPPPG--PPT-----PPR 787
:|||||:
Db 693 HIPSTVQAQSPGHGVYPPSSVAQSQSQSPSSSLTGVSSQPIQHQQOQOQIQOQTAPPQ 752

QY 788 RTSRAPTEPT-PASEATGAPTTPPAPSPSPAPPVVPKKEKEEETAAAPPVE--EGEQK 844
:|||||:
Db 753 QTVQYSLQSTSTSEATTA-----QPVSQPAQVLPQVSAKQLPVSQPVPTIQGEPI 807

QY 845 PPAABE--LAVDTGKABEPVKSETEBAEGPANGKDAEAAEATAEGALKAEKKGSGR 902
:|||||:
Db 808 PVATQPSVVPVHSGAHLFV-----GQPL 831

QY 903 ATTAKSGAQDSDSSATCSADEYDEABGGDKNRLSPRSLLTPTGDPANASPKPLD 962
:|||||:
Db 832 PTPL-----LPQYPVSIPISTPHVSTAQTG-----FSSLPTMA-----AGITQPLL 874

QY 963 LKQLKQRAAIP-----PIQTVKHPPREDAAPTKPAAPPAPPPQNQLQESDAPQPG 1016
:|||||:
Db 875 TLASSATTAATPGVSTVVPVPSQLPTLLQPVTO-----LFSQVHPQLLP-----AVQSMG 923

QY 1017 SSPRGKSRSPAPPADKEAFAAEAKLPGDPPCWTSGLPFPVPPREVIKASPHAPDPSAFS 1076
:|||||:
Db 924 I-----PANL-GQAAEVLSSGD-----VLQGFPPRLPPQY-----PGSN 959

QY 1077 YAPPGHPLPLGLHDTARFVL--PRPTTISNPPPLISSAKHPSV---LEROIGAI SQGMSV 1131
:|||||:
Db 960 IAPSNVASVCIHST---VLSPPMPTEVLATPGYFTVVPVSVESNLLVPNGV--GQV 1014

QY 1132 QLVVPYSEHAKAVPGVPTMGLPLPMDPKLAPFGVKQEQSLSPRGQAQPPPSLVG----- 1186
:|||||:
Db 1015 QVSPQGGSLAQAPTTSQAV-----LESTQGV-----SQVAPFPVAVAPQA 1058

QY 1187 --PTAQEASVLRGALGSLVPGSGSITKGIPTSRVPSDSAITYRGSTTHGTGTPADVLKGIT 1244
:|||||:
Db 1059 TQPTTLTASSV--DSAHSDVASG-MSDG--NENVPSSSG-----RHEGRTT 1098

QY 1245 RIIGEDSPSLDRGREDLSLPGHVIYEGKGHVLVSYEGGMSVTQCSKE----- 1292
```

Db 1099 KRHYKRSVRSRREKTSRPLRLILNVSNGD-----RVBCEQLETHNRKMTKFKFD 1150
Qy 1293 -DGRSSCPHETAPKRTYDMEGRCVRAISSASIEGLMGRAIPPERHSPHLLKEQHII 1351
Db 1151 LDGN-----PEEIIATIMVNDFI-----LATERESFVDQVREII-----EKADEMLSEDSVSV 1198
Qy 1352 RGSITQGIPTRSYBAQEDYLRRREAKLAKRECTGTPPPPPSRDLTEAYKTOALGPLKLKPAH 1411
Db 1199 EPEGDQGL--ESLQKDDYGFSGQKLEGEFKQIPASSM-----PQOIGIPT 1244
Qy 1412 EGLVATYKAGRS--IHEIPRELRHTPELPLAPRPLKBSITQGTPLKVDYDTGASTTGSK 1469
Db 1245 SSLQVQVHSAGRRFIVSPVPSRLRESKVF-----SEIT-----DTVAASTAQS 1289
Qy 1470 -----KHDVRL-----IGSPGRTFPPVPHPLDVMADARALE 1500
Db 1290 PGMNLSHASLSLQAFSELURRAQMTGPNTPNFSHTGTPFPVVPV-----1338
Qy 1501 RACYEELSKRPGTASSSGGSIGARGAP-----VIVPELCKPROSPL--TYEDHGAPFA 1551
Db 1339 ---FLSIAIGVPTTAATAATAPVATSSPNDISTSVIOSEVTVPTTEGIAGVATSGVVS 1395
Qy 1552 GHLPRGSPVTWREPTPLQEGSLSSKASQDRKLTSTPREIA---KSPHSTVTPPEHHPHI 1608
Db 1396 GGL-----PFPVSESPVLSSVVS---ITIPAVVSICTSPSLQVPTSTSEIV 1441
Qy 1609 SPYEHLLRGVGVLDYKSHIPLADPTSIKGIPLDAAAAYLPHLAPNPTPHLYPPY 1668
Db 1442 -----VSSITALYPS-----VTVSATSASAGGSTATPGK-----PPA 1473
Qy 1669 LI-----RGYPDTAALENRQ-----TIINDYITSQOMH 1696
Db 1474 VVSQQAAGSTTVGATLTSVSTTTSPSTASQISLQLSSSTSTPTLAETVVVSAISLQKTS 1533
Qy 1697 HNTATAMA-----QRADMLRGILSPR-----BSSLALNYAAGPRGIID 1733
Db 1534 HSTTGLAFSLAPSSSSSPGAGVSSVISPQGLHPLVIPSIVASTPILPQAAGTSTPL 1593
Qy 1734 LSQVPHLPVLVPPFGTGPATANDRLAYLPTAPQFSSHSSPLSPGCP--THLTKPTTS 1792
Db 1594 LPQVPSIPPLVQPVANPAV---QOTLIHSQFQ-----ALLNPQPHTHCP-----1636
Qy 1793 SSERDRDRDRDREREKSLTSTTVEHAPITWRPGTESGSSGSSGGGSSSRPA 1852
Db 1637 -----EVDSDTQKAPGIDDIKLEB-----KLRSLSSESSSA-----1671
Qy 1853 SHSHAHQSPISPRQDLOQRPSVLHNTGMKGI--ITAVEPSKPTVLRTSTSTSPVRPAA 1911
Db 1672 -----QHASVSLSTS-----LVISTVTPGIPTTAVAPSK--LLTSTT-----S 1708
Qy 1912 TFPPTHCPGLGTLGVPTLMEPVLLPKEAPVAPRPRADTGHAFKAPPARSGLEP 1971
Db 1709 TCLPPTNLPLG-----TVALPVTVPVTCQVSTPVSSTTSGVKP 1747
Qy 1972 ASSPSKSEPRPLVPVSGHATARTPAKNIAPHASDPDPAPPASADPHREKTQS--KP 2030
Db 1748 GTAFSKPPLTKAPVLVPGUTELPAGTLFSEQ-----PFPFGPSL-----TQSQQP 1792
Qy 2031 FSIQELRLSYGHGSSYPGVEFPVSPVSPFSLTHDKGLPHLEELD KSHLEGELRPKQ 2090
Db 1793 LEDLDAQLR-----RTLSPETIITVSAV-----1815
Qy 2091 PGPVKLGGAHLPHLRPLP-----SQPSSSPLLQTAQV-----KHQORVVTLAQHISEVI 2143
Db 1816 -GPVSMAPTA--ITEAGTQPKQSVQVKEGVLATSSGAGVFKMGKGRFQVSA--1866
Qy 2144 TDVTRHHPOQLSAPLAPLAPLSPFFCASCPLDLRRPDSLDLPPP-----DHGAP 2193
Db 1867 --DGAQKCKNKSSEDAKSVHRESSTSSSVLSSSPSTLVKPPENGITIPGISDVPES 1924
Qy 2194 ARGSPHSGGKRSPEPNKTS-----VLGGGEDGIE-----PVSPPGEMTE 2233

Db 1925 AHKTASEAKSDTGQPTKVGRFQVTTTANKVGRFSVSKTEDKITDTTKKEGPFVAGPPFMDL 1984
Qy 2234 PGHSRSAYVLLYRDGQTEPSPRMGSKSPGNTSOP--PAFFSKLTESNSAMVKSQKQIINK 2292
Db 1985 EQAVLPAVVPKKEKP-ELSEPHLN-----CPSSDPEAAFLSRDVGSGSPHSPHQLSSK 2039
Qy 2293 KLANTHNRNEPEYNIQPGTEIFNNPAITGTGLMTYRSQAVQ-----BHASTNMGLEA 2344
Db 2040 SL-----FSQNLQSLSNSFNSVMSNEDIEDLKLRLRLDRKHLKEIQDLQS 2092
Qy 2345 IIRKALMGKYDQWEESPLSANAFNPLNASASUPLAAMPITAADGRSHDTLTPSGGCGKAK 2404
Db 2093 RQKHEIESLTGLKGVPP-----AVIIPPAAPLS---GRRRRPTKS-----KGS 2133
Qy 2405 VSGRPSRKAASP-APGLASGRDRPPSV-----SSVHSEGD 2438
Db 2134 KSSRSSSLGNKSPQLSGNLGQSAASVLPHQTLHPGPN 2172

RESULT 94

Q20007 PRELIMINARY; PRT; 1274 AA.
AC Q20007
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE F35A5.1 protein.
GN F35A5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodirinae; Caenorhabditis.
OX NCBI TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton D., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Leimbach D.;
RA "The sequence of C. elegans cosmid F35A5.";
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U46675; AAB52641.1; -.
DR PIR; T16251; T16251.
DR WormPep; F35A5.1; CE04485.
SQ SEQUENCE 1274 AA; 138065 MW; 07F6B0D292C1799F CRC64;

Query Match 3.1%; Score 404; DB 5; Length 1274;
Best Local Similarity 20.1%; Pred. No. 3.4e-11;
Matches 353; Conservative 167; Mismatches 601; Indels 632; Gaps 87;
Qy 682 KKKKAPAAEEAAPPVVEDEEAEASGVSGNEEVEEAEALHASGNEVPRGCSGPAT 741
Db 14 KKKKPP-----WESVDEEEVEVD-----EETPA-----PSKLEKPPSL 47

Qy 2226 SPPEGMTEPGHRSNAVYLLYRDCQTEP-----SRMGSKSPGNTSOPPAFFSKLTESN 2279
 Db 2573 HSPGAVAGVGTGAFVTWASSWATERAQOQRHSSATGRHQRTVSP-----WRITDLN 2628
 Qy 2280 SAMVSKKQEQINKLINTHRNPEYNIQSGQTEIFNMP--AITG 2321
 Db 2629 SSPVR-----PRWDISQPSPIYHPAPGAITG 2655
 RESULT 97
 Q9VR13
 ID Q9VR13 PRELIMINARY; PRT; 1118 AA.
 AC Q9VR13;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG15635 protein.
 GN CG15635.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkelley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 BA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foeller C., Gabrielista A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon D., Beeson K.Y., Buesam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Flybase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003575; AAF50995.3; -
 DR Flybase; FBgn0031617; CG15635.
 DR GO; GO:0003677; P:DNA binding; IEA.
 DR GO; GO:0008170; P:N-methyltransferase activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR InterPro; IPR002052; N6_Mtase.
 DR PROSITE; PS00092; N6_MTASE; 1.
 SQ SEQUENCE 1118 AA; CAC4F9FB2E26C96D CRC64;
 Query Match 3.08; Score 400.5; DB 5; Length 1118;
 Best Local Similarity 20.5%; Pred. No. 4.3e-11;
 Matches 368; Conservative 109; Mismatches 515; Indels 803; Gaps 88;
 Qy 727 SGNEVPRGECGSP---ATVNNSSDTEIPSPHTEAAKDTGQNGPKPATLGGADGPPPGPP 783
 Db 4 SFDRIKSGSYTPPHFYVDNFTPSAPAP-----PPKP-----RPPPPP 45
 Qy 784 TTPRTSRAPLEPTPASEATCAPPPAPPSPSPAPPVVPKKEKEETAAP-----PV 837
 Db 46 PPPTTTRPTTTTPTTTTPTTTPP--PPPPSAPPPTDP-----ATPGVYPLPI 94
 Qy 838 EEGEQ--KPPAAEELAVDTCKAEPEVKSECTEAEERPAKGDAAEAATAEGALKAEK 895
 Db 95 YGNAPQPGDPPGSHLLS-NSGKHPP-----SSATCSADEVDEAGGDKNLLSPRPSLLTPTGDPR 952
 Qy 896 KEGSGRATTAKSSGAPQDS--SSATCSADEVDEAGGDKNLLSPRPSLLTPTGDPR 952
 Db 120 -----GQMPNPSGFLVPPAICS-----PPYGPEN 144
 Qy 953 ANASQPKLDLKLQKRAAIPPIQV-TKVHEPPR-----EDAATKPA---PPA--- 998
 Db 145 QDGNPPG-----SQTPNSELVPPPAIWNPAYEINQIHPPESQAPNSGFPVPPAIWNP 200
 Qy 999 ---PPPPQNLPQESDAPQCGSSPRGSRSPA-PPADKEAFAAQAQ----- 1040
 Db 201 PYGPNPDGPNPESQTPNPSGFLVPPPAIWNPSVCGTSEDGHPSPESQAPNSGFPVPPAIC 260
 Qy 1041 -----KLFQD--PPCWTSGLEPP-----VPPREVINKASPHADPPSAFSAVPPGHP 1083
 Db 261 YPPFGQNQIGHQSGSHKPP--NSGIYPPSTGWIPTSGSLSGQGHPPG-----SQLPNSG 314
 Qy 1084 LPLGLHTARVLPRTPTISNPPPLISAKHPSVLRIQNAISQMSQVLHVFPSEHAKA 1143
 Db 315 LPPGSIPLLRP-----PNQIGDPP-----GSQMPNSNL----- 341

Job time : 398 secs

This Page Blank (uspto)